



A01

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12Q 1/70, C07K 14/01, C12N 1/21, C12N 15/10, C12N 15/34, C12Q 1/18, C12Q 1/68	A2	(11) International Publication Number: WO 00/32825 (43) International Publication Date: 08 June 2000 (08.06.2000)
(21) International Application Number: PCT/IB99/02040 (22) International Filing Date: 03 December 1999 (03.12.1999) (30) Priority Data: 09/326,144 03 June 1999 (03.06.1999) US 09/407,804 28 September 1999 (28.09.1999) US 09/454,252 02 December 1999 (02.12.1999) US 60/110,992 03 December 1998 (03.12.1998) US 60/157,218 30 September 1999 (30.09.1999) US 60/168,777 01 December 1999 (01.12.1999) US (60) Parent Application or Grant PHAGETECH, INC. [/]; (). PELLETIER, Jerry [/]; (). GROS, Phillippe [/]; (). DUBOW, Michael [/]; (). PELLETIER, Jerry [/]; (). GROS, Phillippe [/]; (). DUBOW, Michael [/]; (). MORROW, Joy, D. ; ().		Published
(54) Title: DEVELOPMENT OF NOVEL ANTI-MICROBIAL AGENTS BASED ON BACTERIOPHAGE GENOMICS (54) Titre: DEVELOPPEMENT DE NOUVEAUX AGENTS ANTIMICROBIENS BASES SUR DES GENOMES DE BACTERIOPHAGES (57) Abstract <p>A method for identifying suitable targets for antibacterial agents based on identifying targets of bacteriophage-encoded proteins is described. Also described are compositions useful in the identification methods and in inhibiting bacterial growth, and methods for preparing and using such compositions.</p> (57) Abrégé <p>L'invention concerne une méthode d'identification de cibles appropriées pour des agents antibactériens. Cette méthode consiste à identifier des cibles de protéines codées par des bactériophages. L'invention concerne également des compositions utiles pour ces méthodes d'identification et pour l'inhibition de la croissance des bactéries. L'invention concerne enfin des méthodes de préparation et d'utilisation de ces compositions.</p>		

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12Q 1/70, 1/68, C12N 15/10, 15/34, 1/21, C07K 14/01, C12Q 1/18		A2	(11) International Publication Number: WO 00/32825 (43) International Publication Date: 8 June 2000 (08.06.00)
(21) International Application Number: PCT/IB99/02040 (22) International Filing Date: 3 December 1999 (03.12.99) (30) Priority Data: 60/110,992 3 December 1998 (03.12.98) US 09/326,144 3 June 1999 (03.06.99) US 09/407,804 28 September 1999 (28.09.99) US 60/157,218 30 September 1999 (30.09.99) US 60/168,777 1 December 1999 (01.12.99) US 09/454,252 2 December 1999 (02.12.99) US (71) Applicant (for all designated States except US): PHAGETECH, INC. [CA/CA]; Place du Parc, Case Postale 387, Montreal H2W 2N9 (CA). (72) Inventors; and (75) Inventors/Applicants (for US only): PELLETIER, Jerry [CA/CA]; 8 Lakeview, Baie D'Urfe, Quebec H9X 3B1 (CA). GROS, Phillippe [CA/CA]; 107 Montrose, St. Lambert, Quebec J4R 1X4 (CA). DUBOW, Michael [CA/CA]; 4901 Coolbrook Avenue, Montreal, Quebec H3X 2K8 (CA).		(74) Agents: MORROW, Joy, D. et al.; Smart & Biggar, 900 - 55 Metcalfe Street, P.O. Box 2999, Station D, Ottawa, Ontario K1P 5Y6 (CA). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(54) Title: DEVELOPMENT OF NOVEL ANTI-MICROBIAL AGENTS BASED ON BACTERIOPHAGE GENOMICS			
(57) Abstract A method for identifying suitable targets for antibacterial agents based on identifying targets of bacteriophage-encoded proteins is described. Also described are compositions useful in the identification methods and in inhibiting bacterial growth, and methods for preparing and using such compositions.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

Description

5

10

15

20

25

30

35

40

45

50

55

DESCRIPTION

Development of Novel Anti-Microbial Agents Based on Bacteriophage Genomics

BACKGROUND OF THE INVENTION

The present invention relates to the field of antibacterial agents and the treatment of infections of animals or other complex organisms by bacteria.

The frequency and spectrum of antibiotic-resistant infections have, in recent years, increased in both the hospital and community. Certain infections have become essentially untreatable and are growing to epidemic proportions in the developing world as well as in institutional settings in the developed world. The staggering spread of antibiotic resistance in pathogenic bacteria has been attributed to microbial genetic characteristics, widespread use of antibiotic drugs, and changes in society that enhance the transmission of drug-resistant organisms. This spread of drug resistant microbes is leading to ever increasing morbidity, mortality and health-care costs.

Ironically, it is the very success of antibiotics, resulting in their widespread use, that has contributed the most to rising numbers of drug resistant bacterial strains. The longer a bacterial strain is exposed to a drug, the more likely it is to acquire resistance. Today, a total of 160 antibiotics, all based on a few basic chemical structures and targeting a small number of metabolic pathways, have found their way to market. Over-prescription of these drugs, as well as the failure of patients to comply with the complete antibiotic regimen, has lead to the rapid emergence of antibiotic resistant strains. Such misuse of prescriptions, careless use of antibiotics in virtually all commercial production of beef and fowl, and changing societal conditions, such as the growth of day-care centers, increased long-term care in hospitals, and increased mobility of the population, has provided an environment where drug-resistant microbes can emerge and spread. Thus, virtually all common infectious bacteria are becoming, or have already become, resistant to one or more groups of antibiotics. Such resistance now reaches all classes of antibiotics currently in use, including: β -lactams, fluoroquinolones, aminoglycosides, macrolide peptides, chloramphenicol, tetracyclines, rifampicin, folate inhibitors, glycopeptides, and mupirocin.

Over the last 45 years bacteria have adapted genetically to avoid the destruction/alteration of the essential pathways that these chemotherapeutic agents

target. Antibiotic resistant bacterial strains are now emerging at a higher rate than the rate at which new antibiotics are being developed. The consequence of this dilemma has been a dramatic increase in the cost of treating infections what would otherwise easily succumb to routine antibiotic therapy. Furthermore, and perhaps most importantly, the emergence of multiple drug resistant pathogenic bacteria has led to a significant increase in morbidity and mortality, particularly in institutional settings.

Most major pharmaceutical companies have on-going drug discovery programs for novel anti-microbials. These are based on screens for small molecule inhibitors (natural products, bacterial culture media, libraries of small molecules, combinatorial chemistry) of crucial metabolic pathways of the micro-organism of interest (e.g., bacteria, fungi, parasites, worms). The screening process is largely for cytotoxic compounds and in most cases is not based on a known mechanism of action of the compounds. Pharmaceutical companies have large programs in this area. Classical drug screening programs are being exhausted and many of these pharmaceutical companies are looking towards rational drug design programs.

Several small to mid-size biotechnology companies as well as large pharmaceutical companies have developed systematic high-throughput sequencing programs to decipher the genetic code of specific micro-organisms of interest. The goal is to identify, through sequencing, unique biochemical pathways or intermediates that are unique to the microorganism. Knowledge of this may, in turn, form the rationale for a drug discovery program based on the mechanism of action of the identified enzymes/proteins. Genome Therapeutics Corp., The Institute for Genome Research, Human Genome Sciences Inc., and other companies have such sequencing programs in place. However, one of the most critical steps in this approach is the ascertainment that the identified proteins and biochemical pathways are 1) non-redundant and essential for bacterial survival, and 2) constitute suitable and accessible targets for drug discovery.

SUMMARY OF THE INVENTION

While animals such as humans are, on occasion, infected by pathogenic bacteria, bacteria also have natural enemies. A number of host-specific viruses, known as bacteriophages or phages, infect and kill bacteria in the natural environment. Such bacteriophages generally have small compact genomes and bacteria are their exclusive hosts. Many known bacteria are host to a large number of bacteriophages that have been described in the literature. During the 1940's - 1960's, phage biology was an area of active research. As a testimony to this, the study of phages which infect and inhibit the enteric bacterium *Escherichia coli* (*E. coli*) contributed much to the early understanding of molecular biology and virology.

As is generally understood, bacteriophage (or phages) are viruses that infect and kill bacteria. They are natural enemies of bacteria and, over the course of evolution, have developed proteins (products of DNA sequences) which enable them to infect a host bacteria, replicate their genetic material, usurp host metabolism, and ultimately kill their host. The scientific literature well documents the fact that many known bacteria have a large number of such bacteriophages (Ackermann and DuBow, 1987) that can infect and kill them (for example, see the ATCC bacteriophage collection at <http://www.atcc.org>).

This invention utilizes the observation that bacteriophages successfully infect and inhibit or kill host bacteria, targeting a variety of normal host metabolic and physiological traits, some of which are shared by all bacteria, pathogenic and nonpathogenic alike. The term "pathogenic" as used herein denotes a contribution to or implication in disease or a morbid state of an infected organism. The invention thus involves identifying and elucidating the molecular mechanisms by which phages interfere with host bacterial metabolism, an objective being to provide novel targets for drug design. Whether the phage blocks bacterial RNA transcription or translation, or attacks other important metabolic pathways, such as cell wall assembly or membrane integrity, the basic blueprint for a phage's bacteria-inhibiting ability is encoded in its genome and can be unlocked using bioinformatics, functional genomics, and proteomics. By these means, the invention utilizes sequence information from the genomics of bacteriophage to identify novel antimicrobials that can be further used to actively and/or prophylactically treat bacterial infection.

Two important components of the invention thus are: i) the identification of bacteria-inhibiting phage open reading frames ("ORF"s) and corresponding products that can be used to develop antibiotics based on amino acid sequence and secondary structural characteristics of the ORF products, and ii) the use of bacteriophages to map

5

4

out essential bacterial target genes and homologs, which can in turn lead to the development of suitable anti-microbial agents. These two avenues represent new and general methods for developing novel antimicrobials.

10

The invention thus concerns the identification of bacteriophage ORFs that supply bacteria-inhibiting functions. In this regard, use of the terms "inhibit", "inhibition", "inhibitory", and "inhibitor" all refer to a function of reducing a biological activity or function. Such reduction in activity or function can, for example, be in connection with a cellular component, *e.g.*, an enzyme, or in connection with a cellular process, *e.g.*, synthesis of a particular protein, or in connection with an overall process of a cell, *e.g.*, cell growth. In reference to bacterial cell growth, for example, an inhibitory effect (*i.e.*, a bacteria-inhibiting effect) may be bacteriocidal (killing of bacterial cells) or bacteriostatic (*i.e.*, stopping or at least slowing bacterial cell growth). The latter slows or prevents cell growth such that fewer cells of the strain are produced relative to uninhibited cells over a given period of time. From a molecular standpoint, such inhibition may equate with a reduction in the level of, or elimination of, the transcription and/or translation of a specific bacterial target(s), or reduction or elimination of activity of a particular target biomolecule.

15

10

20

25

It is particularly advantageous to evaluate a plurality of different phage ORFs for inhibitory activity that may be from one, but is preferably from a plurality of different phage. For example, evaluating ORFs from a number of different phage of the same bacterial host provides at least two advantages. One is that the multiple phages will provide identification of a variety of different targets. Second, it is likely that multiple phage will utilize the same cellular target

30

35

As used herein, the terms "bacteriophage" and "phage" are used interchangeably to refer to a virus which can infect a bacterial strain or a number of different bacterial strains.

40

In the context of this invention, the term "bacteriophage ORF" or "phage ORF" or similar term refers to a nucleotide sequence in or from a bacteriophage. In connection with a particular ORF, the terms refer an open reading frame which has at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity with an ORF from the particular phage identified herein (*e.g.*, with an ORF as identified herein) or to a nucleic acid sequence which has the specified sequence identity percentage with such an ORF sequence.

45

A first aspect of the invention thus provides a method for identifying a bacteriophage nucleic acid coding region encoding a product active on an essential bacterial target by identifying a nucleic acid sequence encoding a gene product which

50

55

5

5

10

provides a bacteria-inhibiting function when the bacteriophage infects a host bacterium, preferably one that is an animal or plant pathogen, more preferably a bird or mammalian pathogen, and most preferably a human pathogen. The bacteriophage is an uncharacterized bacteriophage. Thus, the method excludes, for example, phage λ , ϕ x174, m13 and other *E.coli*-specific bacteriophage that have been studied with respect to gene number and/or function. It also excludes, for example, the nucleic acid coding regions described in Tables 12-14, and in preferred embodiments, excludes the phage in which those regions are naturally located.

15

20

25

30

35

40

45

50

55

In connection with bacteriophage, the term "uncharacterized" means that a certain bacteriophage's genome has not yet been fully identified such that the genes having function involved in inhibiting host cells have not been identified. In particular, phage for which the description of genomic or protein sequence was first provided herein are uncharacterized. Phage sequences for which host bacteria-inhibiting functions have been identified prior to the filing of the present application (or alternatively prior to the present invention) are specifically excluded from the aspects involving utilization of sequences from uncharacterized bacteriophage, except that aspects may involve a plurality of phage where one or more of those phage are uncharacterized and one or more others have been characterized to some extent. A number of different bacteria-inhibiting phage ORFs are indicated in Tables 11-14. The phage ORFs or sequences identified therein are not within the term "uncharacterized; alternatively, in preferred embodiments the phage containing those ORFs are excluded from this term. Further, any additional phage ORFs (or alternatively the phage which contain those ORFs) which have previously been described in the art as bacteria-inhibiting ORFs are expressly excluded; those ORFs or phage are known to those skilled in the art and the exclusion can be made express by specifically naming such ORFs or phage as needed (likewise for uncharacterized targets as described below). For the sake of brevity, such a listing is not expressly presented, as such information is readily available to those skilled in the art.

Stating that an agent or compound is "active on" a particular cellular target, such as the product of a particular gene, means that the target is an important part of a cellular pathway which includes that target and that the agent acts on that pathway. Thus, in some cases the agent may act on a component upstream or downstream of the stated target, including on a regulator of that pathway or a component of that pathway. By "essential", in connection with a gene or gene product, is meant that the host cannot survive without, or is significantly growth compromised, in the absence depletion, or alteration of functional product. An "essential gene" is thus one that encodes a product that is beneficial, or preferably necessary, for cellular growth in

5

6

10

15

20

25

30

35

40

45

50

55

vitro in a medium appropriate for growth of a strain having a wild-type allele corresponding to the particular gene in question. Therefore, if an essential gene is inactivated or inhibited, that cell will grow significantly more slowly, preferably less than 20%, more preferably less than 10%, most preferably less than 5% of the growth rate of the uninhibited wild-type, or not at all, in the growth medium. Preferably, in the absence of activity provided by a product of the gene, the cell will not grow at all or will be non-viable, at least under culture conditions similar to the *in vivo* conditions normally encountered by the bacterial cell during an infection. For example, absence of the biological activity of certain enzymes involved in bacterial cell wall synthesis can result in the lysis of cells under normal osmotic conditions, even though protoplasts can be maintained under controlled osmotic conditions. In the context of the invention, essential genes are generally the preferred targets of antimicrobial agents. Essential genes can encode target molecules directly or can encode a product involved in the production, modification, or maintenance of a target molecule.

A "target" refers to a biomolecule that can be acted on by an exogenous agent, thereby modulating, preferably inhibiting, growth or viability of a cell. In most cases such a target will be a nucleic acid sequence or molecule, or a polypeptide or protein. However, other types of biomolecules can also be targets, *e.g.*, membrane lipids and cell wall structural components.

The term "bacterium" refers to a single bacterial strain, and includes a single cell, and a plurality or population of cells of that strain unless clearly indicated to the contrary. In reference to bacteria or bacteriophage, the term "strain" refers to bacteria or phage having a particular genetic content. The genetic content includes genomic content as well as recombinant vectors. Thus, for example, two otherwise identical bacterial cells would represent different strains if each contained a vector, *e.g.*, a plasmid, with different phage ORF inserts.

In preferred embodiments, the phage is *Staphylococcus aureus* phage 77, 3A, 96, or 44 AHJD, *Enterococcus* sp. phage 182, or *Streptococcus pneumoniae* phage Dp-1.

In preferred embodiments, the phage is selected from. Preferred embodiments involve expressing at least one recombinant phage ORF(s) in a bacterial host followed by inhibition analysis of that host. Inhibition following expression of the phage ORF is indicative that the product of the ORF is active on an essential bacterial target. Such evaluation can be carried out in a variety of different formats, such as on a support matrix such as a solidified medium in a petri dish, or in liquid culture.

5

7

10

15

20

25

30

35

40

45

50

55

Preferably a plurality of phage ORFs are expressed in at least one bacterium. The plurality of phage ORFs can be from one or a plurality of phage. With respect to a single phage or at least one phage in a plurality of phages, the plurality of expressed ORFs preferably represents at least 10%, more preferably at least 20%, 40%, or 60%, still more preferably at least 80% or 90%, and most preferably at least 95% of the ORFs in the phage genome. Preferably, for a plurality of phage, the plurality of expressed ORFs preferably represents at least 10%, more preferably at least 20%, 40%, or 60%, still more preferably at least 80% or 90%, and most preferably at least 95% of the ORFs in the phage genome of each phage. The plurality of phage ORFs can be expressed in a single bacterium, or in a plurality of bacteria where one ORF is expressed in each bacterium, or in a plurality of bacteria where a plurality of ORFs are expressed in at least one or in all of the plurality of bacteria, or combinations of these.

In embodiments of the above aspect (as well as in other aspects herein) in which a plurality of phage are utilized, a plurality of phage have the same bacterial host species; have different bacterial host species; or both. The plurality of phage includes at least two different phage, preferably at least 3, 4, 5, 6, 8, 10, 15, 20, or more different phage. Indeed, more preferably, the plurality of phage will include 50, 75, 100, or more phage. As described herein, the larger number of phage is useful to provide additional target and target evaluation information useful in developing antibacterial agents, for example, by providing identification of a larger range of bacterial targets, and/or providing further indication of the suitability of a particular target (for example, utilization of a target by a number of different unrelated phage can suggest that the target is particularly stable and accessible and effective) and/or can indicate alternate sites on a target which interact with different inhibitors.

Further embodiments involve confirmation of the inhibitor function of the phage ORF, such as by utilizing or incorporating a control(s) designed to confirm the inhibitory nature of the ORF(s) being evaluated. The control can, for example, be provided by expression of an inactive or partially inactive form of the ORF or ORF product, and/or by the absence of expression of the ORF or ORF product in the same or a closely comparable bacterial strain as that used for expression of the test ORF. The reduced level of activity or the absence of active ORF product in the control will thus not provide the inhibition provided by a corresponding inhibitory ORF, or will provide a distinguishably lower level of inhibition. An inactivated or partially inactivated control has a mutation(s), e.g., in the coding region or in flanking regulatory elements, that reduce(s) or eliminate(s) the normal function of the ORF. Thus, the inhibition of a bacterium following expression of a phage ORF is determined by comparison with the effects of expression of an inactivated ORF or the

5

8

10

15

20

25

30

35

40

45

50

55

response of the bacteria in the absence of expression in the same or similar type bacterium. Such determination of inhibition of the bacterium following expression of the ORF is indicative of a bacteria-inhibiting function. These manipulations are routinely understood and accomplished by those of skill in the art using standard techniques. In embodiments utilizing absence of expression of the ORF, the bacteria can, for example, contain an empty vector or a vector which allows expression of an unrelated sequence which is preferably non-inhibitory. Alternatively, the bacteria may have no vector at all. Combinations of such controls or other controls may also be utilized as recognized by those skilled in the art.

In embodiments involving expression of a phage ORF in a bacterial strain, in preferred embodiments that expression is inducible.

By "inducible" is meant that expression is absent or occurs at a low level until the occurrence of an appropriate environmental stimulus provides otherwise. For the present invention such induction is preferably controlled by an artificial environmental change, such as by contacting a bacterial strain population with an inducing compound (*i.e.*, an inducer). However, induction could also occur, for example, in response to build-up of a compound produced by the bacteria in the bacterial culture, *e.g.*, in the medium. As uncontrolled or constitutive expression of inhibitory ORFs can severely compromise bacteria to the point of eradication, such expression is therefore undesirable in many cases because it would prevent effective evaluation of the strain and inhibitor being studied. For example, such uncontrolled expression could prevent any growth of the strain following insertion of a recombinant ORF, thus preventing determination of effective transfection or transformation. A controlled or inducible expression is therefore advantageous and is generally provided through the provision of suitable regulatory elements, *e.g.*, promoter/operator sequences that can be conveniently transcriptionally linked to a coding sequence to be evaluated. In most cases, the vector will also contain sequences suitable for efficient replication of the vector in the same or different host cells and/or sequences allowing selection of cells containing the vector, *i.e.*, "selectable markers." Further, preferred vectors include convenient primer sequences flanking the cloning region from which PCR and/or sequencing may be performed.

As knowledge of the nucleotide sequence of phage ORFs is useful, *e.g.*, for assisting in the identification of phage proteins active against essential bacterial host targets, preferred embodiments involve the sequencing of at least a portion of the phage genome in combination with the above methods. This can be done either before or after or independent of expression and inhibition of the ORF in the bacteria, and provides information on the nature and characteristics of the ORF. Such a portion is

preferably at least 10%, 20%, 40%, 80%, 90%, or 100% of the phage genome. For embodiments in which a plurality of phage are utilized, preferably each phage is sequenced to an extent as just specified.

Such sequencing is preferably accompanied by computer sequence analysis to define and evaluate ORF(s), ORF products, structural motifs or functional properties of ORF products, and/or their genetic control elements. Thus, certain embodiments incorporate computer sequence analyses or nucleic acid and/or amino acid sequences. Further, existing data banks can provide phage sequence and product information which can be utilized for analysis and identification of ORFs in the sequence.

Computer analysis may further employ known homologous sequences from other species that suggest or indicate conserved underlying biochemical function(s) for the inhibitory or potentially inhibitory ORF sequence(s) being evaluated. This can include the sequences of signature motifs of identified classes of inhibitors.

In the context of the phage nucleic acid sequences, e.g., gene sequences, of this invention, the terms "homolog" and "homologous" denote nucleotide sequences from different bacteria or phage strains or species or from other types of organisms that have significantly related nucleotide sequences, and consequently significantly related encoded gene products, preferably having related function. Homologous gene sequences or coding sequences have at least 70% sequence identity (as defined by the maximal base match in a computer-generated alignment of two or more nucleic acid sequences) over at least one sequence window of 48 nucleotides, more preferably at least 80 or 85%, still more preferably at least 90%, and most preferably at least 95%. The polypeptide products of homologous genes have at least 35% amino acid sequence identity over at least one sequence window of 18 amino acid residues, more preferably at least 40%, still more preferably at least 50% or 60%, and most preferably at least 70%, 80%, or 90%. Preferably, the homologous gene product is also a functional homolog, meaning that the homolog will functionally complement one or more biological activities of the product being compared. For nucleotide or amino acid sequence comparisons where a homology is defined by a % sequence identity, the percentage is determined using BLAST programs (with default parameters (Altschul et al., 1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acid Res. 25:3389-3402). Any of a variety of algorithms known in the art which provide comparable results can also be used, preferably using default parameters. Performance characteristics for three different algorithms in homology searching is described in Salamov et al., 1999, "Combining sensitive database searches with multiple intermediates to detect distant

5

10

homologues." *Protein Eng.* 12:95-100. Another exemplary program package is the GCG™ package from the University of Wisconsin.

10

Homologs may also or in addition be characterized by the ability of two complementary nucleic acid strands to hybridize to each other under appropriately stringent conditions. Hybridizations are typically and preferably conducted with probe-length nucleic acid molecules, preferably 20-100 nucleotides in length. Those skilled in the art understand how to estimate and adjust the stringency of hybridization conditions such that sequences having at least a desired level of complementarity will stably hybridize, while those having lower complementarity will not. For examples of hybridization conditions and parameters, see, e.g., Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology, John Wiley & Sons, Secaucus, N.J. Homologs and homologous gene sequences may thus be identified using any nucleic acid sequence of interest, including the phage ORFs and bacterial target genes of the present invention.

25

30

35

40

A typical hybridization, for example, utilizes, besides the labeled probe of interest, a salt solution such as 6xSSC (NaCl and Sodium Citrate base) to stabilize nucleic acid strand interaction, a mild detergent such as 0.5% SDS, together with other typical additives such as Denhardt's solution and salmon sperm DNA. The solution is added to the immobilized sequence to be probed and incubated at suitable temperatures to preferably permit specific binding while minimizing nonspecific binding. The temperature of the incubations and ensuing washes is critical to the success and clarity of the hybridization. Stringent conditions employ relatively higher temperatures, lower salt concentrations, and/or more detergent than do non-stringent conditions. Hybridization temperatures also depend on the length, complementarity level, and nature (ie, "GC content") of the sequences to be tested. Typical stringent hybridizations and washes are conducted at temperatures of at least 40°C, while lower stringency hybridizations and washes are typically conducted at 37°C down to room temperature (~25°C). One of skill in the art is aware that these conditions may vary according to the parameters indicated above, and that certain additives such as formamide and dextran sulphate may also be added to affect the conditions.

45

By "stringent hybridization conditions" is meant hybridization conditions at least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM NaH₂PO₄, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhart's solution at 42°C overnight; washing with 2X SSC, 0.1% SDS at 45°C; and washing with 0.2X SSC, 0.1% SDS at 45°C.

50

55

In sequence comparison analyses, an ORF, or motif, or set of motifs in a bacteriophage sequence can be compared to known inhibitor sequences, *e.g.*, homologous sequences encoding homologous inhibitors of bacterial function. Likewise, the analysis can include comparison with the structure of essential bacterial gene products, as structural similarities can be indicative of similar or replacement biological function. Such analysis can include the identification of a signature, or characteristic motif(s) of an inhibitor or inhibitor class.

Also, the identification of structural motifs in an encoded product, based on nucleotide or amino acid sequence analysis, can be used to infer a biochemical function for the product. A database containing identified structural motifs in a large number of sequences is available for identification of motifs in phage sequences. The database is PROSITE, which is available at www.expasy.ch/cgi-bin/scanprosite. The identification of motifs can, for example, include the identification of signature motifs for a class or classes of inhibitory proteins. Other such databases may also be used.

In aspects and preferred embodiments described herein, in which a bacterium or host bacterium is specified, the bacterium or host bacterium is preferably selected from a pathogenic bacterial species, for example, one selected from Table 1. Preferably, an animal or plant pathogen is used. For animals, preferably the bacterium is a bird or mammalian pathogen, still more preferably a human pathogen.

In aspects and preferred embodiments involving a bacteriophage or sequences from a bacteriophage, one or more bacteriophage are preferably selected from those listed in Table 1. Those exemplary bacteriophage are readily obtained from the indicated sources.

In some cases, it is advantageous to utilize phage with non-pathogenic host bacteria. The genome, structural motif, ORF, homolog, and other analyses described herein can be performed on such phage and bacteria. Such analysis provides useful information and compositions. The results of such analyses can also be utilized in aspects of the present invention to identify homologous ORFs, especially inhibitor ORFs in phage with pathogenic bacterial hosts. Similarly, identification of a target in a non-pathogenic host can be used to identify homologous sequences and targets in pathogenic bacteria, especially in genetically closely related bacteria. Those skilled in the art are familiar with bacterial genetic relationships and with how to determine relatedness based on levels of genomic identity or other measures of nucleotide sequence and/or amino acid sequence similarity, and/or other physical and culture characteristics such as morphology, nutritional requirements, or minimal media-to support growth.

5

12

Also in preferred embodiments, an embodiment of this aspect is combined with an embodiment of the following aspect.

10

A related aspect of the invention provides methods for identifying a target for antibacterial agents by identifying the bacterial target(s) of at least one

5 uncharacterized or untargeted inhibitor protein or RNA from a bacteriophage. Such identification allows the development of antibacterial agents active on such targets. Preferred embodiments for identifying such targets involve the identification of binding of target and phage ORF products to one another. The phage ORF products may be subportions of a larger ORF product that also binds the host target. In 10 preferred embodiments, the phage protein or RNA is from an uncharacterized bacteriophage in Table 1. This aspect preferably includes the identification of a plurality of such targets in one or a plurality of different bacteria, preferably in one or a plurality of bacteria listed in Table 1.

15

20

In preferred embodiments of this aspect and other aspects of this invention 15 involving particular phage ORFs or phage sequences, the ORF is *Staphylococcus aureus* phage 77 ORF 17, 19, 43, 102, 104, or 182 as identified in U.S. application 09/407,804, *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

25

30

As indicated for the above aspect, preferably the method involves the use of a plurality of different phage, and thus a plurality of different phage inhibitors and/or inhibitor ORFs.

35

In addition to uncharacterized phage ORF products, it is also useful to identify the targets of phage ORF products which are known to be inhibitors of host bacteria, 25 but where the target has not been identified. Thus, such inhibitors can likewise be utilized as "untargeted" inhibitor phage ORFs and ORF products, e.g., proteins or RNAs.

40

In the context of inhibitor proteins or RNAs from a phage, the term "uncharacterized" means that a bacteria-inhibiting function for the protein has not 30 previously been identified. Preferably, but not necessarily, the sequence of the protein or the corresponding coding region or ORF was not described in the art before the filing of the present application for patent (or alternatively prior to the present invention). Thus, this term specifically excludes any bacteria-inhibiting phage protein and its associated bacterial target which has been identified as inhibitory before the 45 present invention or alternatively before the filing of the present application, for example those identified in Tables 12-14 or otherwise identified herein. For example, from *E. coli*, phage T7 genes 0.7 and 2.0 target the host RNA polymerase, phage T4 50

55

gp55/gp33 alter the specificity of host RNA polymerase. The T4 *regB* gene product also targets the host translation apparatus. As with the uncharacterized bacteriophage ORFs or bacteriophage above, for such identified proteins, the sequences encoding those proteins are excluded from the uncharacterized inhibitor proteins.

The term "fragment" refers to a portion of a larger molecule or assembly. For proteins, the term "fragment" refers to a molecule which includes at least 5 contiguous amino acids from the reference polypeptide or protein, preferably at least 8, 10, 12, 15, 20, 30, 50 or more contiguous amino acids. In connection with oligo- or polynucleotides, the term "fragment" refers to a molecule which includes at least 15 contiguous nucleotides from a reference polynucleotide, preferably at least 24, 30, 36, 45, 60, 90, 150, or more contiguous nucleotides.

Preferred embodiments involve identification of binding that include methods for distinguishing bound molecules, for example, affinity chromatography, immunoprecipitation, crosslinking, and/or genetic screen methods that permit protein:protein interactions to be monitored. One of skill in the art is familiar with these techniques and common materials utilized (see, e.g., Coligan, J. et al. (eds.) (1995) Current Protocols in Protein Science, John Wiley & Sons, Secaucus, N.J.).

Genetic screening for the identification of protein:protein interactions typically involves the co-introduction of both a chimeric bait nucleic acid sequence (here, the phage ORF to be tested) and a chimcric target nucleic acid sequence that, when co-expressed and having affinity for one another in a host cell, stimulate reporter gene expression to indicate the relationship. A "positive" can thus suggest a potential inhibitory effect in bacteria. This is discussed in further detail in the Detailed Description section below. In this way, new bacterial targets can be identified that are inhibited by specific phage ORF products or derivatives, fragments, mimetics, or other molecules.

Other embodiments involve the identification and/or utilization of mutant targets by virtue of their host's relatively unresponsive nature in the presence of expression of ORFs previously identified as inhibitory to the non-mutant or wild-type strain. Such mutants have the effect of protecting the host from an inhibition that would otherwise occur and indirectly allow identification of the precise responsible target for follow-up studies and anti-microbial development. In certain embodiments, rescue from inhibition occurs under conditions in which a bacterial target or mutant target is highly expressed. This is performed, for example, through coupling of the sequence with regulatory element promoters, e.g., as known in the art, which regulate expression at levels higher than wild-type, e.g., at a level sufficiently higher that the

inhibitor can be competitively bound to the highly expressed target such that the bacterium is detectably less inhibited.

Identification of the bacterial target can involve identification of a phage-specific site of action. This can involve a newly identified target, or a target where the phage site of action differs from the site of action of a previously known antibacterial agent or inhibitor. For example, phage T7 genes 0.7 and 2.0 target the host RNA polymerase, which is also the cellular target for the antibacterial agent, rifampin. To the extent that a phage product is found to act at a different site than previously described inhibitors, aspects of the present invention can utilize those new, phage-specific sites for identification and use of new agents. The site of action can be identified by techniques well-known to those skilled in the art, for example, by mutational analysis, binding competition analysis, and/or other appropriate techniques.

Once a bacterial host target protein or nucleic acid or mutant target sequence has been identified and/or isolated, it too can be conveniently sequenced, sequence analyzed (e.g., by computer), and the underlying gene(s), and corresponding translated product(s) further characterized. Preferred embodiments include such analysis and identification. Preferably such a target has not previously been identified as an appropriate target for antibacterial action.

Certain embodiments include the identification of at least one inhibitory phage ORF or ORF product, e.g., as described for the above aspect, and thus are a combination of the two aspects.

Additionally, the invention provides methods for identifying targets for antibacterial agents by identifying homologs of a bacterial target e.g., *S. aureus*, *Enterococcus faecalis* or other *Enterococci*, and *Streptococcus pneumoniae* of a bacteriophage inhibitory ORF product. Such homologs may be utilized in the various aspects and embodiments described herein as described for the host *Enterococcus* sp. for bacteriophage 182.

Other aspects of the invention provide isolated, purified, or enriched specific phage nucleic acid and amino acid sequences, subsequences, and homologs thereof for phage selected from uncharacterized phage listed in Table 1, preferably from bacteriophage 77, 3A, 96, 44AHJD (*Staphylococcus aureus* host bacterium), Dp-1 (*Streptococcus pneumoniae* host), or 182 (*Enterococcus* host) or other phage listed in Table 1 for those bacteria. For example, such sequences do not include sequences identified in any of Tables 11-14. Nucleotide sequences of this aspect are at least 15 nucleotides in length, preferably at least 18, 21, 24, or 27 nucleotides in length, more preferably at least 30, 50, or 90 nucleotides in length. In certain embodiments, longer

5

15

10

15

20

25

30

35

40

45

50

55

nucleic acids are preferred, for example those of at least 120, 150, 200, 300, 600, 900 or more nucleotides. Such sequences can, for example, be amplification oligonucleotides (e.g., PCR primers), oligonucleotide probes, sequences encoding a portion or all of a phage-encoded protein, or a fragment or all of a phage-encoded protein. In preferred embodiments, the nucleic acid sequence contains a sequence which is within a length range with a lower length as specified above, and an upper length limit which is no more than 50, 60, 70, 80, or 90% of the length of the corresponding full-length ORF. The upper length limit can also be expressed in terms of the number of base pairs of the ORF (coding region). In preferred embodiments, the nucleic acid sequence is from *Staphylococcus aureus* phage 77 ORF 17, 19, 43, 102, 104, or 182 as identified in U.S. application 09/407,804, *S. aureus* phage 44 AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

As it is recognized that alternate codons will encode the same amino acid for most amino acids due to the degeneracy of the genetic code, the sequences of this aspect includes nucleic acid sequences utilizing such alternate codon usage for one or more codons of a coding sequence. For example, all four nucleic acid sequences GCT, GCC, GCA, and GCG encode the amino acid, alanine. Therefore, if for an amino acid there exists an average of three codons, a polypeptide of 100 amino acids in length will, on average, be encoded by 3^{100} , or 5×10^{47} , nucleic acid sequences. Thus, a nucleic acid sequence can be modified (e.g., a nucleic acid sequence from a phage as specified above) to form a second nucleic acid sequence encoding the same polypeptide as encoded by the first nucleic acid sequence using routine procedures and without undue experimentation. Thus, all possible nucleic acid sequences that encode the specified amino acid sequences are also fully described herein, as if all were written out in full, taking into account the codon usage, especially that preferred in the host bacterium. The alternate codon descriptions are available in common textbooks, for example, Stryer, BIOCHEMISTRY 3rd ed., and Lehninger, BIOCHEMISTRY 3rd ed., along with many others. Codon preference tables for various types of organisms are available in the literature. Sequences with alternate codons at one or more sites can also be utilized in the computer-related aspects and embodiments herein. Because of the number of sequence variations involving alternate codon usage, for the sake of brevity, individual sequences are not separately listed herein. Instead the alternate sequences are described by reference to the natural sequence with replacement of one or more (up to all e.g., up to 3, 5, 10, 15, 20, 30, 40, 50, or more) of the degenerate codons with alternate codons from the alternate codon

5

16

10

table (Table 6), or a modified table applicable to a particular organism that has differing codon usage, preferably with selection according to preferred codon usage for the normal host organism or a host organism in which a sequence is intended to be expressed. Those skilled in the art also understand how to alter the alternate codons to be used for expression in organisms where certain codons code differently than shown in the "universal" codon table.

15

For amino acid sequences or polypeptides, sequences contain at least 5 peptide-linked amino acid residues, and preferably at least 6, 7, 10, 15, 20, 30, or 40, amino acids having identical amino acid sequence as the same number of contiguous amino acid residues in a particular phage ORF product. In some cases longer sequences may be preferred, for example, those of at least 50, 60, 70, 80, or 100 amino acids in length. In preferred embodiments, the amino acid sequence contains a sequence which is within a length range with a lower length as specified above, and an upper length limit which is no more than 50, 60, 70, 80, or 90% of the length of the corresponding full-length ORF product. The upper length limit can also be expressed in terms of the number of amino acid residues of the ORF product. In preferred embodiments, the amino acid sequence or polypeptide has bacteria-inhibiting function when expressed or otherwise present in a bacterial cell which is a host for the bacteriophage from which the sequence was derived.

20

25

30

35

By "isolated" in reference to a nucleic acid is meant that a naturally occurring sequence has been removed from its normal cellular (*e.g.*, chromosomal) environment or is synthesized in a non-natural environment (*e.g.*, artificially synthesized). Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but that it is essentially free (about 90-95% pure at least) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

40

45

The term "enriched" means that the specific DNA or RNA sequence constitutes a significantly higher fraction (2-5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in cells from which the sequence was originally taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased.

50

55

The term "significant" is used to indicate that the level of increase is useful to the person making such an increase and an increase relative to other nucleic acids of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The other source DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level, this level should be at least 2-5 fold greater, *e.g.*, in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10⁶-fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The terms "isolated", "enriched", and "purified" as respect nucleic acids, above, may similarly be used to denote the relative purity and abundance of polypeptides (multimers of amino acids joined one to another by α -carboxyl: α -amino group (peptide) bonds). These, too, may be stored in, grown in, screened in, and selected from libraries using biochemical techniques familiar in the art. Such polypeptides may be natural, synthetic or chimeric and may be extracted using any of a variety of methods, such as antibody immunoprecipitation, other "tagging" techniques, conventional chromatography and/or electrophoretic methods. Some of the above utilize the corresponding nucleic acid sequence.

As indicated above, aspects and embodiments of the invention are not limited to entire genes and proteins. The invention also provides and utilizes fragments and portions thereof, preferably those which are "active" in the inhibitory sense described above. Such peptides or oligopeptides and oligo or polynucleotides have preferred lengths as specified above for nucleic acid and amino acid sequences from phage; corresponding recombinant constructs can be made to express the encoded same. Also included are homologous sequences and fragments thereof.

Nucleic acid sequences of the present invention can be isolated using a method similar to those described herein or other methods known to those skilled in the art.

In addition, such nucleic acid sequences can be chemically synthesized by well-known methods. Also, by having particular phage ORFs, e.g., the phage ORFs identified herein (e.g., anti-bacterial ORFs of the present invention, portions thereof, or oligonucleotides derived therefrom as described), other antimicrobial sequences from other bacteriophage sources can be identified and isolated using methods described here or other methods, including methods utilizing nucleic acid hybridization and/or computer-based sequence alignment methods.

The invention also provides bacteriophage antimicrobial DNA segments from other phages based on nucleic acids and sequences hybridizing to the presently identified inhibitory ORF under high stringency conditions or sequences that are highly homologous. The bacteriophage segment from a specific phage, e.g., an antimicrobial DNA segment, can be used to identify a related segment from another unrelated phage based on stringent conditions of hybridization or on being a homolog based on nucleic acid and/or amino acid sequence comparisons. As with identified inhibitory sequences, such homologous coding sequences and products can be used as antimicrobials, to construct active portions or derivatives, to construct peptidomimetics, and to identify bacterial targets.

The nucleotide and amino acid sequences identified herein are believed to be correct, however, certain sequences may contain a small percentage of errors, e.g., 1-5%. In the event that any of the sequences have errors, the corrected sequences can be readily provided by one skilled in the art using routine methods. For example, the nucleotide sequences can be confirmed or corrected by obtaining and culturing the relevant phage, and purifying phage genomic nucleic acids. A region or regions of interest can be amplified, e.g., by PCR from the appropriate genomic template, using primers based on the described sequence. The amplified regions can then be sequenced using any of the available methods (e.g., a dideoxy termination method).

5 This can be done redundantly to provide the corrected sequence or to confirm that the
described sequence is correct. Alternatively, a particular sequence or sequences can
10 be identified and isolated as an insert or inserts in a phage genomic library and
isolated, amplified, and sequenced by standard methods. Confirmation or correction
5 of a nucleotide sequence for a phage gene provides an amino acid sequence of the
encoded product by merely reading off the amino acid sequence according to the
normal codon relationships and/or expressed in a standard expression system and the
15 polypeptide product sequenced by standard techniques. The sequences described
herein thus provide unique identification of the corresponding genes, coding
10 sequences, and other sequences, allowing those sequences to be used in the various
aspects of the present invention.

20 In other aspects, the invention provides recombinant vectors and cells
harboring at least one of the phage ORFs or portion thereof, or bacterial target
sequences described herein. As understood by those skilled in the art, vectors may be
15 provided in different forms, including, for example, plasmids, cosmids, and virus-
based vectors. See, e.g., Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory
25 Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; See also, Ausubel,
F.M. et al. (eds.) (1994) Current Protocols in Molecular Biology. John Wiley & Sons,
Secaucus, N.J.

30 In preferred embodiments, the vectors will be expression vectors, preferably
shuttle vectors that permit cloning, replication, and expression within bacteria. An
"expression vector" is one having regulatory nucleotide sequences containing
transcriptional and translational regulatory information that controls expression of the
nucleotide sequence in a host cell. Preferably the vector is constructed to allow
35 amplification from vector sequences flanking an insert locus. In certain embodiments,
the expression vectors may additionally or alternatively support expression, and/or
replication in animal, plant and/or yeast cells due to the presence of suitable
regulatory sequences, e.g., promoters, enhancers, 3' stabilizing sequences, primer
40 sequences, etc. In preferred embodiments, the promoters are inducible and specific
30 for the system in which expression is desired, e.g., bacteria, animal, plant, or yeast.
The vectors may optionally encode a "tag" sequence or sequences to facilitate protein
purification. Convenient restriction enzyme cloning sites and suitable selective
45 marker(s) are also optionally included. Such selective markers can be, for example,
antibiotic resistance markers or markers which supply an essential nutritive growth
35 factor to an otherwise deficient mutant host, e.g., tryptophan, histidine, or leucine in
the Yeast Two-Hybrid systems described below.

The term "recombinant vector" relates to a single- or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with appropriate restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a desired product can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together. Preferably the vector is an expression vector, e.g., a shuttle expression vector as described above.

By "recombinant cell" is meant a cell possessing introduced or engineered nucleic acid sequences, e.g., as described above. The sequence may be in the form of or part of a vector or may be integrated into the host cell genome. Preferably the cell is a bacterial cell.

In another aspect, the invention also provides methods for identifying and/or screening compounds "active on" at least one bacterial target of a bacteriophage inhibitor protein or RNA. Preferred embodiments involve contacting such a bacterial target or targets (e.g., bacterial target proteins) with a test compound, and determining whether the compound binds to or reduces the level of activity of the bacterial target (e.g., a bacterial target protein). Preferably this is done either *in vivo* (i.e., in a cell-based assay) or *in vitro*, e.g., in a cell-free system under approximately physiological conditions.

The compounds that can be used may be large or small, synthetic or natural, organic or inorganic, proteinaceous or non-proteinaceous. In preferred embodiments, the compound is a peptidomimetic, as described herein, a bacteriophage inhibitor protein or fragment or derivative thereof, preferably an "active portion", or a small molecule.

In preferred embodiments, the bacterial target is a target of a phage ORF identified herein, e.g., *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

In particular embodiments, the methods include the identification of bacterial targets or the site of action of an inhibitor on a bacterial target as described above or otherwise described herein.

In embodiments involving binding assays, preferably binding is to a fragment or portion of a bacterial target protein, where the fragment includes less than 90%, 80%, 70%, 60%, 50%, 40%, or 30% of an intact bacterial target protein. Preferably,

the at least one bacterial target includes a plurality of different targets of bacteriophage inhibitor proteins, preferably a plurality of different targets. The plurality of targets can be in or from a plurality of different bacteria, but preferably is from a single bacterial species.

A "method of screening" refers to a method for evaluating a relevant activity or property of a large plurality of compounds (e.g., a bacteria-inhibiting activity), rather than just one or a few compounds. For example, a method of screening can be used to conveniently test at least 100, more preferably at least 1000, still more preferably at least 10,000, and most preferably at least 100,000 different compounds, or even more.

In the context of this invention, the term "small molecule" refers to compounds having molecular mass of less than 2000 Daltons, preferably less than 1500, still more preferably less than 1000, and most preferably less than 600 Daltons. Preferably but not necessarily, a small molecule is not an oligopeptide.

In a related aspect or in preferred embodiments, the invention provides a method of screening for potential antibacterial agents by determining whether any of a plurality of compounds, preferably a plurality of small molecules, is active on at least one target of a bacteriophage inhibitor protein or RNA. Preferred embodiments include those described for the above aspect, including embodiments which involve determining whether one or more test compounds bind to or reduce the level of activity of a bacterial target, and embodiments which utilize a plurality of different targets as described above.

The identification of bacteria-inhibiting phage ORFs and their encoded products also provides a method for identifying an active portion of such an encoded product. This also provides a method for identifying a potential antibacterial agent by identifying such an active portion of a phage ORF or ORF product. In preferred embodiments, the identification of an active portion involves one or more of mutational analysis, deletion analysis, or analysis of fragments of such products. The method can also include determination of a 3-dimensional structure of an active portion, such as by analysis of crystal diffraction patterns. In further embodiments, the method involves constructing or synthesizing a peptidomimetic compound, where the structure of the peptidomimetic compound corresponds to the structure of the active portion. In this context, "corresponds" means that the peptidomimetic compound structure has sufficient similarities to the structure of the active portion that the peptidomimetic will interact with the same molecule as the phage protein and preferably will elicit at least one cellular response in common which relates to the inhibition of the cell by the phage protein.

In preferred embodiments, the ORF or ORF product is or is derived or obtained from *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014 or product thereof.

The methods for identifying or screening for compounds or agents active on a bacterial target of a phage-encoded inhibitor can also involve identification of a phage-specific site of action on the target.

Preferably in the methods for identifying or screening for compounds active on such a bacterial target, the target is uncharacterized; the target is from an uncharacterized bacterium from Table 1; the site of action is a phage-specific site of action.

Further embodiments include the identification of inhibitor phage ORFs and bacterial targets as in aspects above.

An "active portion" as used herein denotes an epitope, a catalytic or regulatory domain, or a fragment of a bacteriophage inhibitor protein that is responsible for, or a significant factor in, bacterial target inhibition. The active portion preferably may be removed from its contiguous sequences and, in isolation, still effect inhibition.

By "mimetic" is meant a compound structurally and functionally related to a reference compound that can be natural, synthetic, or chimeric. In terms of the present invention, a "peptidomimetic," for example, is a compound that mimics the activity-related aspects of the 3-dimensional structure of a peptide or polypeptide in a non-peptide compound, for example mimics the structure of a peptide or active portion of a phage- or bacterial ORF-encoded polypeptide.

A related aspect provides a method for inhibiting a bacterial cell by contacting the bacterial cell with a compound active on a bacterial target of a bacteriophage inhibitor protein or RNA, where the target was uncharacterized. In preferred embodiments, the compound is such a protein, or a fragment or derivative thereof; a structural mimetic, e.g., a peptidomimetic, of such a protein or fragment; a small molecule; the contacting is performed *in vitro*, the contacting is performed *in vivo* in an infected or at risk organism, e.g., an animal such as a mammal or bird, for example, a human, or other mammal described herein; the bacterium is selected from a genus and/or species listed in Table 1; the bacteriophage inhibitor protein is uncharacterized; the bacteriophage inhibitor protein is from an uncharacterized phage listed in Table 1; the phage inhibitor protein is from one of *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

5

23

10

15

20

25

30

35

40

45

50

55

In the context of targets in this invention, the term "uncharacterized" means that the target was not recognized as an appropriate target for an antibacterial agent prior to the filing of the present application or alternatively prior to the present invention. Such lack of recognition can include, for example, situations where the target and/or a nucleotide sequence encoding the target were unknown, situations where the target was known, but where it had not been identified as an appropriate target or as an essential cellular component, and situations where the target was known as essential but had not been recognized as an appropriate target due to a belief that the target would be inaccessible or otherwise that contacting the cell with a compound active on the target *in vitro* would be ineffective in cellular inhibition, or ineffective in treatment of an infection. Methods described herein utilizing bacterial targets, e.g., for inhibiting bacteria or treating bacterial infections, can also utilize "uncharacterized target sites", meaning that the target has been previously recognized as an appropriate target for an antibacterial agent, but where an agent or inhibitor of the invention is used which acts at a different site than that at which the previously utilized antibacterial agent, i.e., a phage-specific site. Preferably the phage-specific site has different functional characteristics from the previously utilized site. In the context of targets or target sites, the term "phage-specific" indicates that the target or site is utilized by at least one bacteriophage as an inhibitory target and is different from previously identified targets or target sites.

In the context of this invention, the term "bacteriophage inhibitor protein" refers to a protein encoded by a bacteriophage nucleic acid sequence which inhibits bacterial function in a host bacterium. Thus, it is a bacteria-inhibiting phage product.

In the context of this invention, the phrase "contacting the bacterial cell with a compound active on a bacterial target of a bacteriophage inhibitor protein" or equivalent phrases refer to contacting with an isolated, purified, or enriched compound or a composition including such a compound, but specifically does not rely on contacting the bacterial cell with an intact phage which encodes the compound. Preferably no intact phage are involved in the contacting.

Related aspects provide methods for prophylactic or therapeutic treatment of a bacterial infection by administering to an infected, challenged or at risk organism a therapeutically or prophylactically effective amount of a compound active on a target of a bacteriophage inhibitor protein or RNA, or as described for the previous aspect. Preferably the bacterium involved in the infection or risk of infection produces the identified target of the bacteriophage inhibitor protein or alternatively produces a homologous target compound. In preferred embodiments, the host organism is a plant or animal, preferably a mammal or bird, and more preferably, a human or other

5

24

mammal described herein. Preferred embodiments include, without limitation, those as described for the preceding aspect.

10

Compounds useful for the methods of inhibiting, methods of treating, and pharmaceutical compositions can include novel compounds, but can also include compounds which had previously been identified for a purpose other than inhibition of bacteria. Such compounds can be utilized as described and can be included in pharmaceutical compositions.

15

In preferred embodiments of this and other aspects of the invention utilizing bacterial target sequences of a bacteriophage inhibitory ORF product, the target sequence is encoded by a *Staphylococcus* nucleic acid coding sequence, preferably *S. aureus*, a *Streptococcus* nucleic acid coding sequence, preferably *Streptococcus pneumoniae*, or *Enterococcus* nucleic acid coding sequence. Possible target sequences are described herein by reference to sequence source sites.

20

25

The amino acid sequence of a polypeptide target is readily provided by translating the corresponding coding region. For the sake of brevity, the sequences are not reproduced herein. For the sake of brevity, the sequences are described by reference to the GenBank entries instead of being written out in full herein. In cases where the TIGR or GenBank entry for a coding region is not complete, the complete sequence can be readily obtained by routine methods, e.g., by isolating a clone in a phage host genomic library, and sequencing the clone insert to provide the relevant coding region. The boundaries of the coding region can be identified by conventional sequence analysis and/or by expression in a bacterium in which the endogenous copy of the coding region has been inactivated and using subcloning to identify the functional start and stop codons for the coding region.

30

35

In the context of nucleic acid or amino acid sequences of this invention, the term "corresponding" indicates that the sequence is at least 95% identical, preferably at least 97% identical, and more preferably at least 99% identical to a sequence from the specified phage genome, a ribonucleotide equivalent, a degenerate equivalent (utilizing one or more degenerate codons), or a homologous sequence, where the homolog provides functionally equivalent biological function.

40

45

By "treatment" or "treating" is meant administering a compound or pharmaceutical composition for prophylactic and/or therapeutic purposes. The term "prophylactic treatment" refers to treating a patient or animal that is not yet infected but is susceptible to or otherwise at risk of a bacterial infection. The term "therapeutic treatment" refers to administering treatment to a patient already suffering from infection.

50

55

The term "bacterial infection" refers to the invasion of the host organism, animal or plant, by pathogenic bacteria. This includes the excessive growth of bacteria which are normally present in or on the body of the organism, but more generally, a bacterial infection can be any situation in which the presence of a bacterial population(s) is damaging to a host organism. Thus, for example, an organism suffers from a bacterial population when excessive numbers of a bacterial population are present in or on the organism's body, or when the effects of the presence of a bacterial population(s) is damaging to the cells, tissue, or organs of the organism.

The terms "administer", "administering", and "administration" refer to a method of giving a dosage of a compound or composition, *e.g.*, an antibacterial pharmaceutical composition, to an organism. Where the organism is a mammal, the method is, *e.g.*, topical, oral, intravenous, transdermal, intraperitoneal, intramuscular, or intrathecal. The preferred method of administration can vary depending on various factors, *e.g.*, the components of the pharmaceutical composition, the site of the potential or actual bacterial infection, the bacterium involved, and the infection severity.

The term "mammal" has its usual biological meaning referring to any organism of the Class Mammalia of higher vertebrates that nourish their young with milk secreted by mammary glands, *e.g.*, mouse, rat, and, in particular, human, bovine, sheep, swine, dog, and cat.

In the context of treating a bacterial infection a "therapeutically effective amount" or "pharmaceutically effective amount" indicates an amount of an antibacterial agent, *e.g.*, as disclosed for this invention, which has a therapeutic effect. This generally refers to the inhibition, to some extent, of the normal cellular functioning of bacterial cells that renders or contributes to bacterial infection.

The dose of antibacterial agent that is useful as a treatment is a "therapeutically effective amount." Thus, as used herein, a therapeutically effective amount means an amount of an antibacterial agent that produces the desired therapeutic effect as judged by clinical trial results and/or animal models. This amount can be routinely determined by one skilled in the art and will vary depending on several factors, such as the particular bacterial strain involved and the particular antibacterial agent used.

In connection with claims to methods of inhibiting bacteria and therapeutic or prophylactic treatments, "a compound active on a target of a bacteriophage inhibitor protein" or terms of equivalent meaning differ from administration of or contact with an intact phage naturally encoding the full-length inhibitor compound. While an intact phage may conceivably be incorporated in the present methods, the method at

least includes the use of an active compound as specified different from a full length inhibitor protein naturally encoded by a bacteriophage and/or a delivery or contacting method different from administration of or contact with an intact phage encoding the full-length protein. Similarly, pharmaceutical compositions described herein at least include an active compound different from a full-length inhibitor protein naturally encoded by a bacteriophage or such a full-length protein is provided in the composition in a form different from being encoded by an intact phage. Preferably the methods and compositions do not include an intact phage.

In accord with the above aspects, the invention also provides antibacterial agents and compounds active on bacterial targets of bacteriophage inhibitor proteins or RNAs, where the target was uncharacterized as indicated above. As previously indicated, such active compounds include both novel compounds and compounds which had previously been identified for a purpose other than inhibition of bacteria. Such previously identified biologically active compounds can be used in embodiments of the above methods of inhibiting and treating. In preferred embodiments, the targets, bacteriophage, and active compound are as described herein for methods of inhibiting and methods of treating. Preferably the agent or compound is formulated in a pharmaceutical composition which includes a pharmaceutically acceptable carrier, excipient, or diluent. In addition, the invention provides agents, compounds, and pharmaceutical compositions where an active compound is active on an uncharacterized phage-specific site.

In preferred embodiments, the target is as described for embodiments of aspects above.

Likewise, the invention provides a method of making an antibacterial agent. The method involves identifying a target of a bacteriophage inhibitor polypeptide or protein or RNA, screening a plurality of compounds to identify a compound active on the target, and synthesizing the compound in an amount sufficient to provide a therapeutic effect when administered to an organism infected by a bacterium naturally producing the target. In preferred embodiments, the identification of the target and identification of active compounds include steps or methods and/or components as described above (or otherwise herein) for such identification. Likewise, the active compound can be as described above, including fragments and derivatives of phage inhibitor proteins, peptidomimetics, and small molecules. As recognized by those skilled in the art, peptides can be synthesized by expression systems and purified, or can be synthesized artificially. In preferred embodiments the inhibitory phage ORF products is from *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus*

pneumoniae phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

As indicated above, sequence analysis of nucleotide and/or amino acid sequences can beneficially utilize computer analysis. Thus, in additional aspects the invention provides computer-related hardware and media and methods utilizing and incorporating sequence data from uncharacterized phage, e.g., uncharacterized phage listed in Table 1, preferably at least one of *Staphylococcus aureus* phage *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014, or 44 AHJD, *Enterococcus* sp. phage 182, or *Streptococcus pneumoniae* phage Dp-1. In general, such aspects can facilitate the above-described aspects. Various embodiments involve the analysis of genetic sequence and encoded products, as applied to the evaluating bacteriophage inhibitor ORFs and compounds and fragments related thereto. The various sequence analyses, as well as function analyses, can be used separately or in combination, as well as in preceding aspects and embodiments. Use in combination is often advantageous as the additional information allows more efficient prioritizing of phage ORFs for identification of those ORFs that provide bacteria-inhibiting function.

In one aspect, the invention provides a computer-readable device which includes at least one recorded amino acid or nucleotide sequence corresponding to one of the specified phage and a sequence analysis program for analyzing a nucleotide and/or amino acid sequence. The device is arranged such that the sequence information can be retrieved and analyzed using the analysis program. The analysis can identify, for example, homologous sequences or the indicated %s of the phage genome and structural motifs. Preferably the sequence includes at least 1 phage ORF or encoded product, more preferably at least 10%, 20%, 30%, 40%, 50%, 70%, 90%, or 100% of the genomic phage ORFs and/or equivalent cDNA, RNA, or amino acid sequences. Preferably the sequence or sequences in the device are recorded in a medium such as a floppy disk, a computer hard drive, an optical disk, computer random access memory (RAM), or magnetic tape. The program may also be recorded in such medium. The sequences can also include sequences from a plurality of different phage.

In this context, the term "corresponding" indicates that the sequence is at least 95% identical, preferably at least 97% identical, and more preferably at least 99% identical to a sequence from the specified phage genome, a ribonucleotide equivalent, a degenerate equivalent (utilizing one or more degenerate codons), or a homologous sequence, where the homolog provides functionally equivalent biological function.

Similarly, the invention provides a computer analysis system for identifying biologically important portions of a bacteriophage genome. The system includes a data storage medium, e.g., as identified above, which has recorded thereon a nucleotide sequence corresponding to at least a portion of at least one uncharacterized bacteriophage genome, a set of program instructions to allow searching of the sequence or sequences to analyze the sequence, and an output device where the portion includes at least the sequence length as specified in the preceding aspect. The output device is preferably a printer, a video display, or a recording medium. More than one output device may be included. For each of the present computer-related aspects, the bacteriophage are preferably selected from the uncharacterized phage listed in Table 1, more preferably from bacteriophage 77, 3A, 96, 44 AHJD (*S. aureus*), Dp-1 (*Streptococcus pneumoniae*), or 182 (*Enterococcus*).

In keeping with the computer device aspects, the invention also provides a method for identifying or characterizing a bacteriophage ORF by providing a computer-based system for analyzing nucleotide or amino acid sequences, e.g., as describe above. The system includes a data storage medium which has recorded a sequences or sequences as described for the above devices, a set of instructions as in the preceding aspect, and an output device as in the preceding aspect. The method further involves analyzing at least one sequence, and outputting the analysis results to at least one output device.

In preferred embodiments, the analysis identifies a sequence similarity or homology with a sequence or sequences selected from bacterial ORFs encoding products with related biological function; ORFs encoding known inhibitors; and essential bacterial ORFs. Preferably the analysis identifies a probable biological function based on identification of structural elements or characteristic or signature motifs of an encoded product or on sequence similarity or homology. Preferably the uncharacterized bacteriophage is from Table 1, more preferably at least one of bacteriophage 77, 3A, 96, 44 AHJD (*S. aureus*), Dp-1 (*Streptococcus pneumoniae*), or 182 (*Enterococcus*). In preferred embodiments, the method also involves determining at least a portion of the nucleotide sequence of at least one uncharacterized bacteriophage as indicated, and recording that sequence on data storage medium of the computer-based system. In preferred embodiments, the analysis identifies a sequence similarity of homology with a *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

As used in the claims to describe the various inventive aspects and embodiments, "comprising" means including, but not limited to, whatever follows the word "comprising". Thus, use of the term "comprising" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present. By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of". Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present. By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

Further embodiments will be apparent from the following Detailed Description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1A and 1B are flow schematics showing the manipulations used to convert pT0021, an arsenite inducible vector containing the luciferase gene, into pTHA or pTM, two *ars* inducible vectors. Vector pTHA contains BamH I, Sal I, and Hind III cloning sites and a downstream HA epitope tag. Vector pTM contains Bam HI and Hind III cloning sites and no HA epitope tag.

FIGURE 2 is a schematic representation of the cloning steps involved to place the DNA segments of any of ORFs 17/ 19/ 43/ 102/104/182 or other sequences into pTHA to assess inhibitory potential. For subcloning into pTM or pT0021, individual ORFs were amplified by the PCR using oligonucleotides targeting the ATG and stop codons of the ORFs. Using this strategy, Bam HI and Hind III sites were positioned immediately upstream or downstream, respectively of the start and stop codons of each ORF. Following digestion with Bam HI and Hind III, the PCR fragments were subcloned into the same sites of pT0021 or pTM. Clones were verified by PCR and direct sequencing.

FIGURE 3 shows a schematic representation of the functional assays used to characterize the bactericidal and bacteriostatic potential of all predicted ORFs (>33 amino acids) encoded by bacteriophage 77. Fig. 3A) Functional assay on semi-solid support media. Fig. 3B) Functional assay in liquid culture.

FIGURE 4A, B, and C is a bar graph showing the results of a screen in liquid media to assess bacteriostatic or bactericidal activity of 93 predicted ORFs (>33 amino acids) encoded by bacteriophage 77. Growth inhibition assays were performed as detailed in the Detailed Description. The relative growth of *Staphylococcus aureus* transformants harboring a given bacteriophage 77 ORF (identified on the bottom of the graph), in the absence or presence of arsenite, is plotted relative to growth of a *Staphylococcus aureus* transformant containing ORF 5, a non-toxic bacteriophage 77 ORF (which is set at 100%). Each bar represents the average obtained from three *Staph A* transformants grown in duplicate. Bacteriophage 77 ORFs showing significant growth inhibition consist of ORFs 17, 19, 102, 104, and 182.

FIGURE 5 shows a block diagram of major components of a general purpose computer.

FIGURE 6 shows an ORF map for *Streptococcus pneumoniae* bacteriophage Dp-1 showing the ORF identifiers, genomic locations, and orientations of the 85 identified ORFs that were found to have ribosomal binding sites and thus are expected to be expressed.

FIGURE 7 shows a schematic representation of the arsenite-inducible expression system present in a shuttle vector designed to express individual *Streptococcus* bacteriophage Dp-1 ORFs in *Streptococcus*. Various modifications can be readily made to such a vector, or other vectors can be readily constructed to provide inducible expression of ORFs in a particular host bacterium using well-known techniques.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The invention may be more clearly understood from the following description.

The tables will first be briefly described.

Table 1 is a listing of a large number of available bacteriophage that can be readily obtained and used in the present invention.

Table 2 shows the complete nucleotide sequence of the genome of *Staphylococcus aureus* bacteriophage 77.

Table 3 shows a list of all the ORFs from Bacteriophage 77 that were screened in the functional assay to identify those with anti-microbial activity.

Table 4 shows the predicted nucleotide sequence, predicted amino acid sequence, and physiochemical parameters of ORF 17/ 19/ 43/ 102/ 104/ 182]. These include the primary amino acid sequence of the predicted protein, the average molecular weight, amino acid composition, theoretical pI, hydrophobicity map, and predicted secondary structure map.

Table 5 shows homology search results. BLAST analysis was performed with ORFs 17/ 19/ 43/ 102/ 104/ 182 against NCBI non-redundant nucleotide and Swissprot databases. The results of this search indicate that: I) ORF 17 has no significant homology to any gene in the NCBI non-NCBI non-redundant nucleotide database, II) ORF 19 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 59 of bacteriophage phi PVL, III) ORF 43 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 39 of phi PVL, IV) ORF 102 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 38 of phi PVL, V) ORF 104 has no significant homology to any gene in the NCBI non-redundant nucleotide database, VI) ORF 182 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 39 of phi PVL.

Table 6 is a table from Alberts et al., MOLECULAR BIOLOGY OF THE CELL 3rd ed., showing the redundancy of the "universal" genetic code.

Table 7 shows the complete nucleotide sequence of *Staphylococcus aureus* bacteriophage 3A.

Table 8 is a listing of the ORFs identified in *Staphylococcus aureus* bacteriophage 3A.

Table 9 shows the complete nucleotide sequence of *Staphylococcus aureus* bacteriophage 96.

Table 10 is a listing of the ORFs identified in *Staphylococcus aureus* bacteriophage 96.

Table 11 is a listing of sequences deposited in the NCBI public database (GeneBank) for bacteriophage listed in Table 1.

Table 12 is a listing of phage which encode a known lysis function, including the identified lysis gene.

Table 13 is a listing of bacteriophage which encode holin genes, where holin genes encode proteins which form pores and eventually enable other enzymes to kill the host bacterium.

Table 14 is a listing of bacteriophage which encode kil genes.

Table 15 is a list of *Staphylococcus aureus* sequences identified by accession number which may include sequences from genes coding for target sequences for the phage 77-encoded antimicrobial proteins or peptides. The sequences were obtained by searching GenBank for listings.

Table 16 shows the nucleotide sequence of the genome of *Staphylococcus aureus* phage 44 AHJD.

Table 17 lists and shows the sequence position of the 73 ORFs predicted to be encoded by *Staphylococcus aureus* bacteriophage 44 AHJD that are greater than 33 amino acids.

Table 18 shows the ORF sequences and putative amino acid sequences for the *Staphylococcus aureus* bacteriophage 44AHJD ORFs greater than 33 amino acids.

Table 19 shows the similarities in sequence identified between predicted *Staphylococcus aureus* bacteriophage 44 AHJD ORFs and sequences present in public databases.

Table 20 shows the homology alignments between predicted *Staphylococcus aureus* bacteriophage 44AHJD ORFs and the corresponding protein sequences present in public sequence databases.

Table 21 shows the complete nucleotide sequence of the genome of *Enterococcus* bacteriophage 182.

Table 22 lists and shows the sequence position of the 80 ORFs identified in bacteriophage 182 and that are greater than 33 amino acids.

Table 23 shows the nucleotide and predicted amino acid sequence of all 80 ORFs identified in bacteriophage 182.

Table 24 shows the similarities identified to date in sequence between *Enterococcus* phage 182 ORFs greater than 33 amino acids and sequences present in public sequence databases.

Table 25 shows the predicted amino acid sequence as well as the predicted secondary structures map for two *Enterococcus* bacteriophage 182 ORFs.

Table 26 shows the homology alignments between predicted *Enterococcus* bacteriophage 182 ORFs and the corresponding protein sequences present in public sequence databases.

Table 27 list *Enterococcus* sequences listed in GenBank providing possible *Enterococcal* target sequences for inhibitory *Enterococcus* bacteriophage 182 ORFs and other compounds with antibacterial activity.

Table 28 shows the complete nucleotide sequence of the genome of *Streptococcus* bacteriophage Dp-1.

Table 29 lists and shows sequence position of the 273 ORFs identified in *Pneumococcal* bacteriophage Dp-1 that are greater than 33 amino acids, 85 of which are predicted to be expressed in Dp-1 as having a ribosomal binding site. That set of 85 ORFs is shown in the attached drawings.

Table 30 shows the nucleotide and predicted amino acid sequence of all 273 ORFs identified in bacteriophage Dp-1 that are identified as being expressed.

Table 31 shows the similarities identified in sequence between *Streptococcus* phage Dp-1 ORFs greater than 33 amino acids and sequences present in public sequence databases.

Table 32 shows the 4731 bp sequence of Dp-1 published by Sheehan et al., 1997).

Table 33 lists *Streptococcus pneumoniae* sequences listed in GenBank providing possible target sequences for inhibitory *Streptococcus pneumoniae* bacteriophage Dp-1 ORFs and other compounds with antibacterial activity

Background:

As indicated above, the present invention is concerned, in part, with the use of bacteriophage coding sequences and the encoded polypeptides or RNA transcripts to identify bacterial targets for potential new antibacterial agents. Thus, the invention concerns the selection of relevant bacteria. Particularly relevant bacteria are those which are pathogens of a complex organism such as an animal, e.g., mammals,

reptiles, and birds, and plants. Examples include *Staphylococcus aureus*, *Enterococcus* species, and *Streptococcus pneumoniae*. However, the invention can be applied to any bacterium (whether pathogenic or not) for which bacteriophage are available or which are found to have cellular components closely homologous to components targeted by phage of another bacterium.

Thus, the invention also concerns the bacteriophage which can infect a selected bacterium. Identification of ORFs or products from the phage which inhibit the host bacterium both provides an inhibitor compound and allows identification of the bacterial target affected by the phage-encoded inhibitor. Such targets are thus identified as potential targets for development of other antibacterial agents or inhibitors and the use of those targets to inhibit those bacteria. As indicated above, even if such a target is not initially identified in a particular bacterium, such a target can still be identified if a homologous target is identified in another bacterium. Usually, but not necessarily, such another bacterium would be a genetically closely related bacterium. Indeed, in some cases, a phage-encoded inhibitor can also inhibit such a homologous bacterial cellular component.

The demonstration that bacteriophage have adapted to inhibiting a host bacterium by acting on a particular cellular component or target provides a strong indication that that component is an appropriate target for developing and using antibacterial agents, *e.g.*, in therapeutic treatments. Thus, the present invention provides additional guidance over mere identification of bacterial essential genes, as the present invention also provides an indication of accessibility of the target to an inhibitor, and an indication that the target is sufficiently stable over time (*e.g.*, not subject to high rates of mutation) as phage acting on that target were able to develop and persist. Thus, the present invention identifies a subset of essential cellular components which are particularly likely to be appropriate targets for development of antibacterial agents.

The invention also, therefore, concerns the development or identification of inhibitors of bacteria, in addition to the phage-encoded inhibitory proteins (or RNA transcripts), which are active on the targets of bacteriophage-encoded inhibitors. As described herein, such inhibitors can be of a variety of different types, but are preferably small molecules.

The following description provides preferred methods for use in the various aspects of the invention. However, as those skilled in the art will readily recognize, other approaches can be used to obtain and process relevant information. Thus, the invention is not limited to the specifically described methods. In addition, the following description provides a set of steps in a particular order. That series of steps

5

35

10

describes the overall development involved in the present invention. However, it is clear that individual steps or portions of steps may be usefully practiced separately, and, further, that certain steps may be performed in a different order or even bypassed if appropriate information is already available or is provided by other sources or methods.

5

Selecting and Growing Phage, and Isolating DNA

15

Conceptually, the first step involves selecting bacterial hosts of interest. Preferably, but not necessarily, such hosts will be pathogens of clinical importance. Alternatively, because bacteria all share certain fundamental metabolic and structural features, these features can be targeted for study in one strain, for example a nonpathogenic one, and extrapolated to similarly succeed in pathogenic ones. Nonpathogenic strains may also exhibit initial advantages in being not only less dangerous, but also, for example, in having better growth and culturing characteristics and/or better developed molecular biology techniques and reagents. Consequently, advantageously the invention provides the ability target virtually any bacteria, but preferably pathogenic bacteria, with antimicrobial compounds designed and/or developed using bacteriophage inhibitory proteins and peptides from phage with non-pathogenic and/or pathogenic hosts.

25

30

We have selected *Staphylococcus aureus*, *Streptococcus pneumoniae*, various *Enterococci*, and *Pseudomonas aeruginosa* as initial exemplary pathogens. These bacteria are a major cause of morbidity and mortality in hospital-based infections, and the appearance of antibiotics resistance in all three organisms makes it increasingly difficult to treat benign infections involving these organisms. Such infections can include, for example, otitis media, sinusitis, and skin, and airway infections (Neu, H.C. (1992). *Science* 257, 1064-1073). However, the approach described below is clearly applicable to any human bacterial pathogens including but not restricted to *Mycobacterium tuberculosis*, *Nisseria gonorrhoeae*, *Haemophilus influenza*, *Acinobacter*, *Escherichia coli*, *Shigella dysenteria*, *Streptococcus pyogenes*, *Helicobacter pylori*, and *Mycoplasma* species. This invention can also be applied to the discovery of anti-bacterial compounds directed against pathogens of animals other than humans, for example, sheep, cattle, swine, dogs, cats, birds, and reptiles. Similarly, the invention is not limited to animals, but also applies to plants and plant pathogens.

35

40

45

In general, the bacteria are grown according to standard methodologies employed in the art, including solid, semi-solid or liquid culturing, which procedures can be found in or extrapolated from standard sources such as Maloy, S.R., Stewart,

50

55

5 V.J., and Taylor, R.K. Genetic Analysis of Pathogenic Bacteria (1996) Cold Spring
Harbor Laboratory Press, or Maniatis, T. et al. (1989) Molecular Cloning: A
10 Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; or
Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley &
5 Sons, Secaucus, N.J. Culture conditions are selected which are adapted to the
particular bacterium generally using culture conditions known in the art as
appropriate, or adaptations of those conditions.

15 Nucleic acids within these bacteria can be routinely extracted through common
procedures such as described in the above-referenced manuals and as generally known
10 to those skilled in the art. Those nucleic acid stocks can then be used to practice the
other inventive aspects described below.

20 Selection and Growth of Bacteriophage, and Isolation of DNA

The second step involves assembling a group of bacteriophages (phage
15 collection) for one or more of the targeted bacterial hosts. While the invention can be
utilized with a single bacteriophage for a pathogen or other bacterium, it is preferable
25 to utilize a plurality of phage for each bacterium, as comparisons between a plurality
of such phage provides useful additional information. Non-limiting examples of
phage and sources for some of the above-mentioned pathogenic bacteria are found in
20 Table 1. The criteria used to select such phages is that they are infectious for the
microbe targeted, and replicate in, lyse, or otherwise inhibit growth of the bacterium
in a measurable fashion. These phages can be very different from one another
(representing different families), as judged by criteria such as morphology (head, tail,
plate, etc.), and similarity of genome nucleotide sequence (cross-hybridization). Since
35 such diverse bacteriophages are expected to block bacterial host metabolism and
ultimately inhibit by a variety of mechanisms, their combined study will lead to the
identification of different mechanisms by which the phages independently inhibit
bacterial targets. Examples include degradation of host DNA (Parson K.A., and
40 Snustad, D.P. (1975). *J. Virol.* 15, 221-444) and inhibition of host RNA transcription
(Severinova, E., Severinov, K. and Darst, S.A. (1998). *J.Mol. Biol.* 279, 9-18). This,
30 in turn, yields novel information on phage proteins that can inhibit the targeted
microbe. As explained below, this 1) forms the basis of novel drug discovery efforts
based on knowledge of the primary amino acid sequence of the phage inhibitor
45 protein (e.g., peptide fragments or peptidomimetics) and/or 2) leads to the
identification of bacterial biochemical pathways, the proteins of which are essential or
35 significant for survival of the targeted microbe, and which enzymatic steps or

chemical reactions can be targeted by classical drug discovery methods using molecular inhibitors, for example, small molecule inhibitors.

Bacteriophage are generally either of two types, lytic or filamentous, meaning they either outright destroy their host and seek out new hosts after replication, or else continuously propagate and extrude progeny phage from the same host without destroying it. Regardless of the phage life cycle and type, preferred embodiments incorporate phage which impede cell growth in measurable fashion and preferably stop cell growth. To this end, lytic phage are preferred, although certain nonlytic species may also suffice, *e.g.*, if sufficiently bacteriostatic.

Various procedures that are commonly understood by those of skill in the art can be routinely employed to grow, isolate, and purify phage. Such procedures are exemplified by those found in such common laboratory aids such as Maloy, S.R., Stewart, V.J., and Taylor, R.K. Genetic Analysis of Pathogenic Bacteria (1996) Cold Spring Harbor Laboratory Press; Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; and Ausubel, F.M. et al. (eds.) (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J. The techniques generally involve the culturing of infected bacterial cells that are lysed naturally and/or chemically assisted, for example, by the use of an organic solvent such as chloroform that destroys the host cells thereby liberating the phage within. Following this, the cellular debris is centrifuged away from the supernatant containing the phage particles, and the phage then subsequently and selectively precipitated out of the supernatant using various methods usually employing the use of alcohols and/or other chemical compounds such as polyethylene glycol (PEG). The resulting phage can be further purified using various density gradient/centrifugation methodologies. The resulting phage are then chemically lysed, thereby releasing their nucleic acids that can be conveniently precipitated out of the supernatant to yield a viral nucleic acid supply of the phage of interest.

Exemplary bacteriophage are indicated in Table 1, along with sources where those phage may be obtained.

Exemplary bacteria include the reference bacteria for the identified bacteriophage, available from the same sources.

Characterizing Bacteriophage Genomes for ORFs

The third step involves systematically characterizing the genetic information contained in the phage genome. Within this genetic information is the sequence of all RNAs and proteins encoded by the phage, including those that are essential or

instrumental in inhibiting their host. This characterization is preferably done in a systematic fashion. For example, this can be done by first isolating high molecular weight genomic DNA from the phage using standard bacterial lysis methods, followed by phage purification using density gradient ultracentrifugation, and extraction of nucleic acid from the purified phage preparation. The high molecular weight DNA is then analyzed to determine its size and to evaluate a proper strategy for its sequencing. The DNA is broken down into smaller size fragments by sonication or partial digestion with frequently cutting restriction enzymes such as Sau3A to yield predominantly 1 to 2 kilobase length DNA, which DNA can then be resolved by gel electrophoresis followed by extraction from the gel.

The ends of the fragments are enzymatically treated to render them suitable for cloning and the pools of fragments are cloned in a bacterial plasmid to generate a library of the phage genome. Several hundred of these random DNA fragments contained in the plasmid vector are isolated as clones after introduction into an appropriate bacterium, usually *Escherichia coli*. They are then individually expanded in culture and the DNA from each individual clone is purified. The nucleotide sequences of the inserts of these clones are determined by standard automated or manual methods, using oligonucleotide primers located on either side of the cloning site to direct polymerase mediated sequencing (e.g., the Sanger sequencing method or a modification of that method). Other sequencing methods can also be used.

The sequence of individual clones is then deposited in a computer, and specific software programs (for example, Sequencher™, Gene Codes Corp.) are used to look for overlap between the various sequences, resulting in ordering of contig sequences and ultimately providing the complete sequence of the entire bacteriophage genome (one such example is given in Table 2 for *Staphylococcus aureus* bacteriophage 77; others are also provided herein). This complete nucleotide sequence is preferably determined with a redundancy of at least 3- to 5-fold (number of independent sequencing events covering the same region) in order to minimize sequencing errors.

Preferably, the bacterial strain used as a phage host should not possess any other innate plasmids, transposons, or other phage or incompatible sequences that would complicate or otherwise make the various manipulations and analyses more difficult.

Commercially available computer software programs are used to translate the nucleotide sequence of the phage to identify all protein sequences encoded by the phage (hereafter called open reading frames or ORFs). (Customized software can clearly also be used.) As phages are known to transcribe their genome into RNA from

both strands, in both directions, and sometimes in more than one frame for the same sequence, this exercise is done for both strands and in all six possible reading frames. As evolutionary constraints have forced the phage to conserve all of its vital protein sequences in as small a genome as possible, it is straightforward to identify all the proteins encoded by the phage by simple examination of the 6 translation frames of the genome. Once these ORFs are identified, they are cataloged into a phage proteome database (Table 3 lists ORFs identified from phage 77; ORF lists are also provided for other exemplary phage). This analysis is preferably performed for each phage under study. The process of ORF identification can be varied depending on the desired results. For example, the minimum length for the putative encoded polypeptide can be varied, and/or putative coding regions that have an associated Shine-Dalgarno sequence can be selected. In the case of phage 77 ORFs, such parameter adjustment was performed and resulted in the identification of ORFs as listed herein. Different parameters had resulted in the identification of the ORFs listed in the preceding U.S. Provisional Application 60/110,992, filed December 3, 1998, which is hereby incorporated by reference in its entirety.

Exemplary phage 77 ORFs identified in that provisional application and as identified herein are shown in the following table:

ORF ID from 60/110,992	Genomic position	a.a. size	Start codon	ORF ID from 241/190	Genomic position	a.a. size	Start codon
77ORF016	2369-24024	251	TTG	77ORF017	23269-23982	237	ATG
77ORF019	39845-40501	218	ATA	77ORF019	39851-40501	216	ATG
77ORF050	29268-29564	98	ATG	77ORF182	29268-29564	98	ATG
77ORF050	29268-29564	98	ATG	77ORF043	29304-29564	86	ATG
77ORF067	34312-34551	79	CTG	77ORF104	34393-34551	52	ATG
77ORF146	29051-29212	53	ATG	77ORF102	29051-29212	53	ATG

Identifying and Characterizing Inhibitory Phage ORFs

The fourth step entails identifying the phage protein or proteins or RNA transcripts that have the ability to inhibit their bacterial hosts. This can be accomplished, for example, by either or both of two non-mutually exclusive methods. The first method makes use of bioinformatics. Over the past few years, a large amount of nucleotide sequence information and corresponding translated products have become available through large genome sequencing projects for a variety of organisms including mammals, insects, plants, unicellular eukaryotes (yeast and fungi), as well as several bacterial genomes such as *E. coli*, *Mycobacterium tuberculosis*, *Bacillus subtilis*, *Staphylococcus aureus* and many others. Such sequences have been deposited in public databases (for example, non-redundant

5

40

10

sequence database at GenBank and SwissProt protein sequence database) (<http://www.ncbi.nlm.nih.gov>) and can be freely accessed to compare any specific query sequence to those present in such databases. For example, GenBank contains over 1.6 billion nucleotides corresponding to 2.3 million sequence records. Several computer programs and servers (e.g., TBLASTN) have been created to allow the rapid identification of homology between any given sequence from one organism to that of another present in such databases, and such programs are public and available free of charge.

15

20

In addition, it has been well established that basic biochemical pathways can be conserved in very distant organisms (for example bacteria and man), and that the proteins performing the various enzymatic steps in these pathways are themselves conserved at the amino acid sequence level. Thus, proteins performing similar functions (e.g. DNA repair, RNA transcription, RNA translation) have frequently preserved key structural signatures, identifiable by similarities across regions of proteins (domains and motifs). The antimicrobials of the present invention will preferably target features and targets that are highly characteristic or conserved in microbes, and not higher organisms.

25

30

Most genomes encode individual proteins or groups of proteins that can be assembled into protein families that have been evolutionarily conserved. Therefore, similarity between a new query sequence and that of a member of a protein family (reference sequences from public databases) can immediately suggest a biochemical function for the novel query sequence, which in our case is a phage ORF.

35

40

The sequence homology between individual members of evolutionarily distant members of a protein family is usually not randomly distributed along the entire length of the sequence but is often clustered into "motifs" and "domains". These correspond to key three-dimensional folds that form key catalytic and/or regulatory structures that perform key biochemical function(s) for the group of proteins. Commercially available computer software programs can identify such motifs in a new query sequence, again providing functional information for the query sequence. Such structural and functional motifs have also been derived from the combined analysis of primary sequence databases (protein sequences) and protein structure databases (X-ray crystallography, nuclear magnetic resonance) using so-called "threading" methods (Rost B, and Sander C. (1996) *Ann. Rev. Biophys. Biomol. Struct.* 25, 113-136).

45

50

Such motifs and folds are themselves deposited in public databases which can be directly accessed (for example, SwissProt database; 3D-ALI at EMBL, Heidelberg; PROSITE). This basic exercise leads to a structural homology map in which each of

55

the phage ORFs has been probed for such similarities, and where initial structural and functional hits are identified (selected examples of sequence homologies detected between individual ORFs from the genome of *Staphylococcus aureus* bacteriophage 77 and sequences deposited in public databases are shown in Table 5 for ORFs 17/19/43/102/104/182).

This analysis can point out phage proteins with similarity to proteins from other phages (such as those for *E. coli*) playing an important role in the basic biochemical pathways of the phage (such as DNA replication, RNA transcription, tRNAs, coat protein and assembly). Selected examples of such proteins include integrase and capsid protein. Therefore, this analysis enables identification and elimination of non-essential ORFs as candidates for an inhibitor function, as well as the identification of (potentially) useful ones.

In addition, this analysis can point out specific ORFs as possible inhibitor ORFs. For example these ORFs may encode proteins or enzymes that alter bacterial cell structure, metabolism or physiology, and ultimately viability. Examples of such proteins present in the genome of *Staphylococcus aureus* bacteriophage 77 include orf14 (deoxyuridine triphosphatase from bacteriophage T5), and orf15 (sialidase). (These ORF identifications are as listed in provisional application 60/110,992.) Other examples include ORFs 9 and 12 of *S. aureus* phage 44 AHJD, which encode the putative lysis functions found in many bacteriophages – a “holin” and an “amidase”.

In addition, it is well known that bacterial and eukaryotic viruses can usurp pathways from their host in order to use them to their advantage in blocking host cellular pathways upon infection. The phage can achieve this by 1) directly producing an inhibitor of a key host pathway (e.g. T7 gene 0.5 and 2), 2) directly producing a novel activity (e.g. T4 DNA polymerase), and 3) altering concentrations of cell components by producing similar functions (e.g. T4 transfer RNAs). The identification of sequence similarity between phage ORFs and bacterial host genome sequences will be highly indicative of such a mechanism. (Selected examples of such homologies are listed in Figure 4 of the provisional application 60/110,992 and include orf4 (homologous to autolysin), orf20 (hypothetical protein from *Staphylococcus aureus*) and orf29 (hypothetical protein from *Staphylococcus aureus*)). These ORFs can be analyzed by a standard biochemical approach to directly test their inhibitor functions (e.g., as described below).

Alternatively, a homology search may reveal that a given phage ORF is related to a protein present in the databases having an activity known to be inhibitory, (e.g., inhibitor of host RNA polymerase by *E. coli* bacteriophage T7. Such a finding would implicate the phage ORF product in a related activity. This will also suggest that a

new antimicrobial could be derived by a mimetic approach (e.g., peptidomimetic) imitating this function or by a small molecule inhibitor to the bacterial target of the phage ORF, or any steps in the relevant host metabolic pathway, e.g., high throughput screening of small molecule libraries. Selected examples of such similarity between ORFs of *Staphylococcus aureus* bacteriophage 77 and proteins with inhibitor functions for bacterial hosts are listed in Figure 4 of the provisional application 60/110,992. These include orf9 (similar to bacteriophage P1 *ki1A* function), and orf4 (autolysin of *Staphylococcus aureus*, amidase enzymatic activity).

A reason for the biochemical study of individual ORFs for inhibitor function is that their expression or overexpression will block cellular pathways of the host, ultimately leading to arrest and/or inhibition of host metabolism. In addition, such ORFs can alter host metabolism in different ways, including modification of pathogenicity. Therefore, individual ORFs identified above are expressed, preferably overexpressed, in the host and the effect of this expression or overexpression on host metabolism and viability is measured. This approach can be systematically applied to every ORF of the phage, if necessary, and does not rely on the absolute identification of candidate ORFs by bioinformatics. Individual ORFs are resynthesized from the phage genomic DNA, e.g., by the polymerase chain reaction (PCR), preferably using oligonucleotide primers flanking the ORF on either side. These single ORFs are preferably engineered so that they contain appropriate cloning sites at their extremities to allow their introduction into a new bacterial expression plasmid, allowing propagation in a standard bacterial host such as *E. coli*, but containing the necessary information for plasmid replication in the target microbe such as *S. aureus* (hereafter referred to as shuttle vector). Shuttle vectors and their use are well known in the art.

Such shuttle vectors preferably also contain regulatory sequences that allow inducible expression of the introduced ORF. As the candidate ORF may encode an inhibitor function that will eliminate the host, it is beneficial that it not be expressed prior to testing for activity. Thus, screening for such sequences when expressed in a constitutive fashion is less likely to be successful when the inhibitor is lethal. In the exemplary inducible system presented in Figure 1A, 1B, 2, and 7, regulatory sequences from the *ars* operon of *S. aureus* are used to direct individual ORF expression in *S. aureus* (or other bacteria in which the *ars* system is functional). The *ars* operon encodes a series of proteins which normally mediate the extrusion of arsenite and other trivalent oxyanions from the cells when they are exposed to such toxic substances in their environment. The operon encoding this detoxifying mechanism is normally silent and only induced when arsenite-related compounds are

present. (Tauriainen, S. et al. (1997) *App. Env. Microb.*, Vol. 63, No. 11, p. 4456-4461.)

Therefore, individual phage ORFs can be expressed in *S. aureus* in an inducible fashion by adding to the culture medium non-toxic arsenite concentrations during the growth of individual *S. aureus* clones expressing such individual phage ORFs. Toxicity of the phage inhibitor ORF for the host is monitored by reduction or arrest of growth under induction conditions, as measured by optical density in liquid culture or after plating the induced cultures on solid medium. Subsequently, interference of the phage ORF with the host biochemical pathways ultimately leading to reduced or arrested host metabolism can be measured by pulse-chase experiments using radiolabeled precursors of either DNA replication, RNA transcription, or protein synthesis. Similar constructs can be made and used for other bacteria using well-known techniques.

Those skilled in the art are familiar with a variety of other inducible systems which can also be used for the controlled expression of phage ORFs, including, for example, lactose (see *e.g.*, Stratagene's LacSwitch™II system; La Jolla, CA) and tetracycline-based systems (see, *e.g.* Clontech's Tet On/Tet Off™ system; Palo Alto, CA). The arsenite-inducible system described is further depicted in Figures 1, 2 and 7.

The selection or construction of shuttle vectors and the selection and use of inducible systems are well known and thus other shuttle vectors appropriate for other bacteria can be readily provided by those skilled in the art, *e.g.*, for use in other bacterial species.

Standard methodologies for expressing proteins from constructs, and isolating and manipulating those proteins, for example in cross-linking and affinity chromatography studies, may be found in various commonly available and known laboratory manuals. See, *e.g.*, Current Protocols in Protein Science, John Wiley & Sons, Secaucus, N.J., and Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.

It has been found that certain phage or other viruses inhibit host cells, at least in part, by producing an antisense RNA which binds to and inhibits translation from a bacterial RNA sequence. Thus, in the case of potentially inhibitor RNA transcripts encoded by the phage genome, a strong indicator of a possible inhibitory function is provided by the identification of phage sequence which is the identical to or fully complementary (or with only a small percentage of mismatch, *e.g.*, <10%, preferably less than 5%, most preferably less than 3%, to a bacterial sequence. This approach is convenient in the case of bacteria that have been essentially completely sequenced, as the comparison can be performed by computer using public database information.

5

44

10

The inhibitory effect of the transcript can be confirmed using expression of the phage sequence in a host bacterium. If needed, such inhibitory can also be tested by transfecting the cells with a vector that will transcribe the phage sequence to form RNA in such manner that the RNA produced will not be translated into a polypeptide. Inhibition under such conditions provides a strong indication that the inhibition is due to the transcript rather than to an encoded polypeptide.

15

In an alternative, the expression of an ORF in a host bacterium is found to be inhibitory, but the inhibition is found to be due to an RNA product of the genomic coding region. For antisense inhibition, the sequence of the bacterial target nucleic acid sequence can be identified by inspection of the phage sequence, and the full sequence of the relevant coding region for the bacterial product can be found from a database of the bacterial genomic sequence or can be isolated by standard techniques (e.g., a clone in a genomic library can be isolated which contains the full bacterial ORF, and then sequenced).

20

25

In either case, the identification of a target which is inhibited by an RNA transcript produced by a phage provides both the possible inhibition of bacteria naturally containing the same target nucleic acid sequence, as well as the ability to use the target sequence in screening for other types of compounds which will act directly on the target nucleic acid sequence or on a polypeptide product expressed or regulated, at least in part, by the target of the inhibitory phage RNA.

30

35

In some cases it will be found that the target of an inhibitory phage RNA or protein has previously been found to be a target of an inhibitory phage RNA or protein has previously been found to be a target for an antibacterial agent. In such cases, the phage inhibitor can still provide useful information if it is found that the phage-encoded product acts at a different site than the previously identified antibacterial agent or inhibitor, i.e., acts at a phage-specific site. For many targets, action at a different site provides highly beneficial characteristics and/or information. For example, an alternate site of inhibitor action can at least partially overcome a resistance mechanism in a bacterium. As an illustration, in many cases, resistance is due, in large part, to altered binding characteristics of the immediate target to the antibacterial agent. The altered binding is due to a structural change which prevents or destabilizes the binding. However, the structural change is frequently quite local, so that compounds which bind at different local sites will be unaffected or affected to a much lesser degree. Indeed, in some cases the local sites will be on a different molecule and so may be completely unaffected by the local structural change creating resistance to the original agent(s). An example of resistance due to altered binding is

40

45

50

55

5

45

provided by methicillin-resistant *Staphylococcus aureus*, in which the resistance is due to an altered penicillin-binding protein.

10

In other cases, a new site of action can have improved accessibility as compared to a site acted on by a previously identified agent. This can, for example, assist in allowing effective treatment at lower doses, or in allowing access by a larger range of types of compounds, potentially allowing identification of more potential active agents.

15

Another advantage is that the structural characteristics of a different site of action will lead to identification and/or development of inhibitors with different structures and different pharmacological parameter. This can allow a greater range of possibilities when selecting an antibacterial agent.

20

Yet further, different sites often produce different inhibitory characteristics in the target organism. This is commonly the case for multi-domain target proteins. Thus, inhibition targeting an alternate site can produce more efficacious action, e.g., faster killing, slower development of resistance, lower numbers of surviving cells, and different secondary effects (for example, different nutrient utilization).

25

Staphylococcus aureus phage 77

30

As indicated above, the present invention is concerned, in part, with the use of bacteriophage 77 coding sequences and the encoded polypeptides or RNA transcripts to identify bacterial targets for potential new antibacterial agents.

35

As described, phage 77 ORFs 17, 19, 43, 102, 104, and 182 have been found to have bacteria inhibiting function. Identification of ORFs 17, 19, 43, 102, 104, and 182 and products from the phage which inhibit the host bacterium both provides an inhibitor compound and allows identification of the bacterial target affected by the phage-encoded inhibitor. Such a target is thus identified as a potential target for development of other antibacterial agents or inhibitors and the use of those targets to inhibit those bacteria. As indicated above, even if such a target is not initially identified in a particular bacterium, such a target can still be identified if a homologous target is identified in another bacterium. Usually, but not necessarily, such another bacterium would be a genetically closely related bacterium. Indeed, in some cases, an inhibitor encoded by phage 77 ORF 17, 19, 43, 102, 104, or 182 can also inhibit such a homologous bacterial cellular component.

40

45

50

Possible bacterial target sequences are described herein by reference to sequence source sites. In preferred embodiments, the sequence encoding the target corresponds

55

to a *S. aureus* nucleic acid sequence available from numerous sources including *S. aureus* sequences deposited in GenBank, *S. aureus* sequences found in European Patent Application No. 97100110.7 to Human Genome Sciences, Inc. filed January 7, 1997, *S. aureus* sequences available from TIGR at <http://www.tigr.org/tdb/mdb/mdb.html>, and *S. aureus* sequences available from the Oklahoma University *S. aureus* sequencing project at the following URL: http://www.genome.ou.edu/staph_new.html. Such possible targets are particularly applicable to *S. aureus* phages 77, 3A, 96, and 44 AHJD.

The amino acid sequence of a polypeptide target is readily provided by translating the corresponding coding region. For the sake of brevity, the sequences are not reproduced herein. Also, in preferred embodiments, a target sequence corresponds to a *S. aureus* coding sequence corresponding to a sequence listed in Table 15 herein. The listing in Table 15 describes *S. aureus* sequences currently listed with GenBank. Again, for the sake of brevity, the sequences are described by reference to the database accession numbers instead of being written out in full herein. In cases where an entry for a coding region is not complete, the complete sequence can be readily obtained by routine methods, e.g., by isolating a clone in a phage host *S. aureus* genomic library, and sequencing the clone insert to provide the relevant coding region. The boundaries of the coding region can be identified by conventional sequence analysis and/or by expression in a bacterium in which the endogenous copy of the coding region has been inactivated and using subcloning to identify the functional start and stop codons for the coding region.

Staphylococcus aureus phage 44 AHJD

The present invention also can utilize the identification of naturally occurring DNA sequence elements within *Staphylococcus aureus* bacteriophage 44AHJD which encode proteins with antimicrobial activity.

Such identification can utilize bioinformatics identification of specific proteins (ORFs) utilized by *Staphylococcus aureus* bacteriophage 44AHJD during the viral life cycle, resulting in a slowing or arrest of growth of the bacterial host, or in death, of the *Staphylococcus aureus* host including lysis of the infected bacteria. Thus, some of the bacteriophage 44AHJD DNA sequences encoding these proteins (ORFs) are predicted to encode antimicrobial functions. Information derived from these DNA sequences and translated ORFs can, in turn, be utilized to develop inhibitory compounds by peptidomimetics that can also function as antimicrobials. In addition, the identification of the host bacterial proteins that are targeted and inhibited by the

antimicrobial bacteriophage ORFs can themselves provide novel targets for drug discovery.

The methodology described above is used to identify and characterize DNA sequences from *Staphylococcus* sp. bacteriophage 44 AHJD that have antimicrobial activity. As described in the Examples, the *Staphylococcus aureus* propagating strain (PS 44A), obtained from the Felix d'Herelle Reference Centre (#HER 1101), was used as a host to propagate its phage 44AHJD, also obtained from the Felix d'Herelle Reference Centre (#HER 101). By sequencing, we found that bacteriophage 44AHJD consists of 16,668 bp (Table 16) predicted to encode 73 ORFs greater than 33 amino acids (Tables 17 & 18). Computational analysis of the predicted protein products of *Staphylococcus aureus* bacteriophage 44AHJD identified homologs in public sequence databases as listed in Table 19 and 20, along with the accompanying list of related proteins.

From this analysis, it is apparent that 3 genes (ORF 3, 7, and 8) are related to structural proteins found in other bacteriophages. These include genes predicted to encode a tail protein (ORF 3), an upper collar/connector protein of the phage virion (ORF 7), and a lower collar protein (ORF 8). Bioinformatics has also identified one gene whose product is likely involved in phage DNA synthesis. One gene (ORF 1) shows significant homology to DNA polymerases of a number of bacteriophages, bacteria and fungi, and the product of this gene is likely responsible for replicating the genetic material of bacteriophage 44AHJD. ORF 2 encodes a protein with homology to the *dinC* gene of *Bacillus subtilis* that encodes a protein involved in teichoic acid biosynthesis. Teichoic acid is a polyphosphate polymer found in some, but not all, Gram positive organisms (and not in Gram negative organisms), where it is attached to the peptidoglycan layer. The phage protein may thus be involved in the synthesis of this material for incorporation into the cell wall, allowing enhanced lysis by the phage lysis enzymes or, as many enzymes can function in "reverse reactions", may be involved in its degradation allowing for penetration of the peptidoglycan and phage genome entry into the cell following adsorption. The similarity between *Staphylococcus aureus* bacteriophage 44AHJD and *E. coli* phage T7 indicates that they may share similar mechanisms of replication and growth. Both phages belong to the Podoviridae Family of bacteriophages and are members of the "T7-like" Genus of this Family (Ackermann and DuBow; VIth ICTV Report).

Two genes, ORF 9 and 12, were identified with the potential to encode antimicrobial protein products. The homology alignments are shown in Tables 19 and 20. The predicted product of ORF 9 is related to a class of genes which encodes lysozyme-like functions, enzymes which cleave linkages in the mucopolysaccharide cell wall structure of a variety of micro-organisms, including that from the *Staphylococcus aureus* bacteriophage Twort. ORF 12 of *Staphylococcus aureus* bacteriophage 44AHJD shows homology to a set of lysis proteins from several bacteriophages. These lysis proteins are also referred to as holins, and represent phage-encoded lysis functions required for transit of the phage murein hydrolases (lysozyme) to the periplasm, where it can digest the cell wall and thus lyse the bacterium.

Thus, in particular embodiments, the present invention provides a nucleic acid sequence isolated from *Staphylococcus aureus* bacteriophage 44AHJD comprising at least a portion of one of the genes described above with antimicrobial activity. For example, ORF 1 encodes a DNA polymerase function. This polymerase may utilize host-derived accessory proteins for its activity when replicating the phage template, sequestering such proteins from use by the bacterial polymerase, resulting in inhibition of DNA replication, cell division, and cell growth. Alternatively, ORF 9 directly encodes a polypeptide with antimicrobial activity. ORF 9 is predicted to encode an amidase, a protein known to act as a cell wall degrading enzyme. ORF 12 likely encodes a holin function required for transit of the phage amidase (gene 9 product) to the periplasm. When this type of gene product from Bacillus phage phi 29 (gene 14), was cloned in *Escherichia coli*, cell death ensued (Steiner et al., 1993). Thus, production of proteins from Bacillus phage phi 29 gene 14 in *E. coli* resulted in cell death, whereas production of protein from Bacillus phage phi 29 gene 14 concomitantly with the phi 29 lysozyme or unrelated murein-degrading enzymes led to lysis, suggesting that membrane-bound protein 14 induces a nonspecific lesion in the cytoplasmic membrane (Steiner et al., 1993).

The present invention also provides the use of the *Staphylococcus* bacteriophage 44 AHJD antimicrobial ORFs or ORF products as pharmacological agents, either wholly or in part and derivatives, as well as the use of corresponding peptidomimetics, developed from amino acid or nucleotide sequence knowledge derived from *Staphylococcus* bacteriophage 44 AHJD killer ORFs.

Enterococcus phage 182

Bacteriophage 182 was obtained from the Felix D'Herelle phage collection (Ste. Foy, Quebec) and infects *Enterococcus* sp. Group D. The genome of *Enterococcus* bacteriophage 182 consists of 17,833 bp (Table 21) and is predicted to encode 80 ORFs greater than 33 amino acids (Tables 22 and 23). Computational analysis of the predicted protein products of *Enterococcus* bacteriophage 182 was performed in order to identify protein products related to those deposited in public databases. Bacteriophage 182 protein products which detected sequences with significant sequence similarity in public databases are listed in Table 24 and 26, along with the accompanying list of related proteins.

From this analysis, it is apparent that 5 genes (ORF 001, 004, 007, 009, and 011) are related to structural proteins of several *Bacillus* phages – *Bacillus* bacteriophage PZA, phi-29, and B103. These include genes predicted to encode a tail protein (ORF 001), a head protein (ORF 004), and upper collar protein (ORF 007), a lower collar protein (ORF 009), and a pre-neck appendage protein (ORF 011). Two gene products are predicted to encode genes which direct phage morphogenesis – these are ORF 005 and 019.

Bioinformatics has also identified three genes whose products are likely involved in phage DNA synthesis. One gene, ORF 002 shows significant homology to DNA polymerases of a number of bacteriophages, and the product of this gene is likely responsible for replicating the genetic material of bacteriophage 182. ORF 006 encodes a protein with homology to the encapsidation proteins of several other bacteriophages, including *Bacillus* phage phi-29 (P11014), PZA (P07541), and B103 (X99260) and *Streptococcus* phage CP-1 (Z47794). These gene products catalyze the *in vivo* and *in vitro* genome-encapsidation reaction (Garvey et al., 1985). Proteins involved in genome packaging have been shown to have additional activities that affect biochemical reactions in other phages and their hosts. For example, the coat protein of the RNA bacteriophage MS2 interacts with viral RNA to translationally repress replicase synthesis (Pickett and Peabody, 1993). This protein-RNA interaction also plays a role in genome encapsidation, enveloping a single copy of the viral genome in a protein shell composed of many molecules of coat protein. In addition, the bacteriophage λ terminase enzyme can be lethal to *E. coli* when expressed,

5

50

10

15

20

25

30

35

40

45

50

55

suggesting cleavage of packaging sites in the bacterial chromosome. Also present within bacteriophage 182 is a gene, ORF 010, that encodes a protein that is related to the terminal proteins of *Bacillus* phage Nf (P06812), *Bacillus* phage GA-1 (X96987) and *Bacillus* phage B103 (X99260). DNA terminal proteins are linked to the 5' ends of both strands of the genome and are essential for DNA replication playing a role in initial priming of DNA replication. The similarity between *Enterococcus* bacteriophage 182 and *Bacillus* phages phi-29, PZA, and B103 indicates that they may share similar mechanisms of replication and growth. Protein-primed DNA replication is a well described phenomenon, and in the phi-29-like phages, the ends of the DNA serve as origins and termini of replication (Gutiérrez et al., 1986; Yoshikawa et al., 1985).

There is also a gene (ORF 015) that encodes a protein showing homology to an early protein product of *Bacillus* bacteriophage PZA and the single-strand nucleic acid binding protein of bacteriophage B103.

Two genes, ORF 008 and 014, were identified with the potential to encode anti-microbial protein products. The homology alignments are shown in Tables 24 & 26 and biochemical features of the predicted polypeptides shown in Table 25. The predicted product of ORF 008 is related to a class of genes which encodes lysozyme-like functions, enzymes which cleave linkages in the mucopolysaccharide cell wall structure of a variety of micro-organisms. ORF 014 of *Enterococcus* 182 shows homology to a set of lysis proteins from *Bacillus* bacteriophage phi-29, PZA, and B103. These lysis proteins are also referred to as holins and represent phage encoded lysis functions required for transit of the phage murein hydrolases (lysozyme) to the periplasm, where it can digest the outer cell wall and thus lyse the bacterium.

Thus, the present invention provides a nucleic acid sequence obtained from *Enterococcus* bacteriophage 182 comprising at least a portion of a phage 182 ORF, preferably an inhibitory ORF, and more preferably at least a portion of one of the genes described above with anti-microbial activity. For example, ORF 002 encodes a DNA polymerase function. This polymerase may utilize host-derived accessory proteins for its activity when replicating the phage template, sequestering such proteins from use by the bacterial polymerase, resulting in inhibition of DNA replication, cell division, and cell growth. Alternatively, ORFs 008 or 014 directly encode polypeptides with anti-microbial activity. ORF 008 is predicted to encode an

autolytic lysozyme, a protein known to have anti-microbial activity (Martin *et al.*, 1998). ORF 014 likely encodes a holin function required for transit of the phage murein hydrolases to the periplasm. When the related product from *Bacillus* phage phi 29 (gene 14), was cloned in *Escherichia coli*, cell death ensued (Steiner *et al.*, 1993).

Thus, production of proteins from *Bacillus* phage phi 29 gene 14 in *E. coli* resulted in cell death, whereas production of protein from *Bacillus* phage phi 29 gene 14 concomitantly with the phi 29 lysozyme or unrelated murein-degrading enzymes led to lysis, suggesting that membrane-bound protein 14 induces a nonspecific lesion in the cytoplasmic membrane (Steiner *et al.*, 1993).

The present invention also provides the use of the *Enterococcus* bacteriophage 182 anti-microbial ORFs as pharmacological agents, either wholly or in part and derivatives, as well as the use of corresponding peptidomimetics, developed from amino acid or nucleotide sequence knowledge derived from *Enterococcus* bacteriophage 182 killer ORFs. This can be done where the structure of the peptidomimetic compound corresponds to the structure of the active portion of a product of an ORF. In this analysis, the peptide backbone is transformed into a carbon-based hydrophobic structure that can retain cytostatic or cytotoxic activity for the bacterium. This is done by standard medicinal chemistry methods, measuring growth inhibition of the various molecules in liquid cultures or on solid medium. These mimetics also represent lead compounds for the development of novel antibiotics. In this context, "corresponds" means that the peptidomimetic compound structure has sufficient similarities to the structure of the active portion of a product of one of the *Enterococcus* ORFs listed, that the peptidomimetic will interact with the same molecule as the product of the ORF, and preferably will elicit at least one cellular response in common which relates to the inhibition of the cell by the phage protein.

To validate the identity of an ORF as a killer ORF, it is preferably expressed in the host or other test bacterial organism and the effect of this expression on bacterial growth and replication is assessed. Therefore, all individual ORFs identified herein, e.g., those identified above, can be expressed, preferably overexpressed, in a suitable host bacterium e.g., a host *Enterococcus* and the effect of this expression or overexpression on host metabolism and viability can be measured.

Individual ORFs can be resynthesized from the phage genomic DNA by the polymerase chain reaction (PCR) using oligonucleotide primers flanking the ORF on

5 either side. Those skilled in the art are familiar with the design and synthesis of
appropriate primer sequences. These single ORFs are preferably engineered so that
10 they contain appropriate cloning sites at their extremities to allow their introduction
into a new bacterial expression plasmid, allowing propagation in a standard bacterial
5 host such as *E. coli*, but containing the necessary information for plasmid replication
in the target microbe, *Enterococcus* sp. (hereafter referred to as a shuttle vector).

15 This shuttle vector also preferably contains regulatory sequences that allow
inducible expression of the introduced ORF. As the candidate ORF may encode a
killer function that will eliminate the host, it is highly advantageous that it not be
10 expressed (or at least not expressed at a substantial level) prior to testing for activity;
20 thus screening for such sequences in a constitutive fashion is less likely to be
successful (lethality). In an example presented in Fig. 7, regulatory sequences from
the *ars* operon are used to direct individual ORF expression in *Enterococcus*. The *ars*
operon encodes a series of proteins which normally mediate the extrusion of arsenite
25 and several other trivalent oxyanions from the cells when they are exposed to such
toxic substances in their environment. The operon encoding this detoxifying
mechanism is normally silent and only induced when arsenite-related compounds are
present.

30 Therefore, individual phage ORFs can be expressed in *Enterococcus* or other
20 suitable host in an inducible fashion by adding to the culture medium non-toxic
arsenite concentrations during the growth of individual *Enterococcus* (or other host
35 cells) clones expressing such individual phage ORFs. Toxicity of the phage killer
ORF for the host is monitored by reduction or arrest of growth under induction
conditions, as measured by optical density in liquid culture or after plating the
25 induced cultures on solid medium. Subsequently, interference of the phage ORF with
40 the host biochemical pathways ultimately leading to reducing or arresting host
metabolism can be measured by pulse chase experiments using radiolabeled
precursors of either DNA replication, RNA transcription, or protein synthesis.

45 Of course, other inducible regulatory sequences (e.g., promoters, operators,
30 etc.) may be used (e.g., systems using positive induction of expression or systems
using release of repression). A variety of such systems are known to those skilled in
the art and can be utilized in the present invention.

5

10

15

20

25

30

35

40

45

50

55

Nucleic acid sequences of the present invention can be isolated using a method similar to those described herein or other methods known to those skilled in the art. In addition, such nucleic acid sequences can be chemically synthesized by well-known methods. Having the phage 182 ORFs, e.g., anti-bacterial ORFs of the present invention, portions thereof, or oligonucleotides derived therefrom as described, other anti-microbial sequences from other bacteriophage sources can be identified and isolated using methods described here or other methods, including methods utilizing nucleic acid hybridization and/or computer-based sequence alignment methods.

The invention also provides bacteriophage anti-microbial DNA segments from other phages based on nucleic acids and sequences hybridizing to the presently identified inhibitory ORF under high stringency conditions or sequences which are highly homologous. The bacteriophage anti-microbial DNA segment from bacteriophage 182 can be used to identify a related segment from another unrelated phage based on stringent conditions of hybridization or on being a homolog based on nucleic acid and/or amino acid sequence comparisons. As with the phage 182 inhibitory sequences, such homologous coding sequences and products can be used as antimicrobials, to construct active portions or derivatives, to construct peptidomimetics, and to identify bacterial targets.

Enterococcus sequences are listed in Table 27 by accession number, providing identification of possible targets of *Enterococcus* phage inhibitory ORF products, e.g., from phage 182.

Streptococcus pneumoniae

As indicated in the Summary above, the present invention is concerned with the use of *Streptococcus* sp. bacteriophage Dp-1 coding sequences and the encoded polypeptides or RNA transcripts to identify bacterial targets for potential new antibacterial agents.

Streptococcus pneumoniae is an important cause of community-acquired pneumonia and a major cause of otitis media, sinusitis, and meningitis in children and adults. In Spain and other Mediterranean countries, the majority of *S. pneumoniae* are relatively resistant to penicillin (Klugman, 1990; Fenoll et al., 1991; Jorgensen et al., 1990). These strains also have decreased susceptibility to broad-spectrum cephalosporins, which are frequently used in the empiric treatment of meningitis and

5

54

other serious invasive bacterial infections. High-level resistance of pneumococci has been encountered in Hungary where 70% of children who were colonized with *S. pneumoniae* carried penicillin resistant strains that were also resistant to tetracycline, erythromycin, trimethoprim/sulfamethoxazole, and 30% resistant to chloramphenicol (Neu, 1992). The resistance of pneumococci to macrolides such as erythromycin averages 20-25% in France, ~20% in Japan, and <10% in Spain (Neu, 1992).

The antimicrobial susceptibilities and distribution of serotypes of the 42 isolates of *S. pneumoniae* in southern Taiwan from invasive infections have been recently determined (Hseuh et al., 1996). Resistance rates among these isolates were: erythromycin, 61.9%; clindamycin, 47.6%; chloramphenicol, 19%; and tetracycline, 73.8%. Resistance to three or more classes of antibiotics was found in 33.3% of the isolates. Bacteremic pneumonia and primary bacteremia accounted for 64.3% of the infections and mortality was 42.6%. Given the severity of these infections despite adequate antibiotic therapy, there is clearly a need for introduction of new therapeutic options to prevent mortality due to invasive *S. pneumoniae* infections.

Pneumococcal phages belong to four families and they present a great variety in morphology, including lytic and temperate phages (for a review, see Garcia et al., 1997). Examples of lytic phages are Cp-1 and Dp-1, whereas examples of temperate phages are HB-3, EJ-1, and HB-746. The complete nucleotide sequence and functional organization of Cp-1 has been reported (Martin et al., 1996). Cp-1 has a 19,345 bp double-stranded DNA genome, with a terminal protein covalently linked to its 5' ends, that replicates by a protein primed mechanism. The phage contains 29 ORFs, 23 on one strand and 6 on the opposite. When these predicted proteins were compared to sequences compiled in GenBank EMBL databases, to ORFs showed significant similarity to proteins of bacteriophage 29 that infects *B. subtilis* (Martin et al., 1996). The similar proteins corresponded to those involved in DNA replication (terminal protein and DNA polymerase), structural and morphogenic proteins (major head, collar, connector, tail, and encapsidation proteins), and proteins involved in lysis function (holin and lysozyme). In its strategy of lysis, the holin gene product inserts itself into the cell membrane, allowing access of the lysozyme to the peptidoglycan. Expression of the Cp-1 holin protein in *E. coli* results in cell death after 2-hours of induction, but did not lead to lysis (Garcia et al., 1997). Cells harboring a plasmid construction with holin and lysozyme genes together did lyse after induction and the

55

5 viability loss was similar to that of the culture expressing holin alone. Cloning of
these lytic genes in *S. pneumoniae* showed that both genes had the same effect as in *E.*
10 *coli*. That is, holin itself did not lyse the culture but the viability loss was noticeable,
whereas both holin and lysozyme together were capable of lysing M31, an amidase
5 deleted mutant (Garcia et al., 1997).

15 Recently, a small portion (~4 kbp) of a second *S. pneumoniae* phage, Dp-1,
has been sequenced (Sheehan et al., 1997). This portion contains the genes coding for
the lytic system (Sheehan et al., 1997) and shows a modular organization similar to
that described for Cp-1. However, in this case, a single chimeric protein appears to be
10 made in which the N-terminal domain is highly similar to that of the murein hydrolase
coded by a gene found in the phage BK5-T that infects *Lactococcus lactis*, and the C-
20 terminal domain is homologous to holins. Thus, both functions appear to have been
combined in a novel chimeric protein.

25 Bacteriophage Dp-1 was obtained from Dr. P. Garcia (Departamento de
15 Microbiologia Molecular, Centro de Departamento de Investigaciones Biologicas,
Consejo Superior de Investigaciones Cientificas, Velazquez, Madrid, Spain). We
found that Dp-1 has a double-stranded DNA genome of 56,506 bp, predicted to
30 encode 85 ORFs greater than 33 amino acids and with upstream Shine-Dalgarno
motifs for translation initiation (Tables 28 & 30, and Fig. 6). Computational analysis
20 of the predicted protein products of *Streptococcus* bacteriophage Dp-1 protein
products, which detected homologs in public databases, are listed in Table 31, along
35 with the accompanying list of related proteins.

From this analysis, it is apparent that several predicted genes of Dp-1 encode
polypeptides that are related to structural proteins. ORFs 001, 002, 004, and 030 are
25 predicted to encode tail proteins, minor structural proteins, and minor capsid proteins
(Table 31). We also note the identification of several gene products that are likely
40 involved in DNA synthesis. These include ORF 3 which encodes DNA polymerase,
ORF 8 which encodes a SWI/SNF helicase-related protein, ORF 10 encodes a protein
showing homology to recA, and ORF 13 encodes a dnaZX-like ORF.

45 In *E. coli*, RapA encodes an RNA polymerase (RNAP)-associated protein with
30 ATPase activity and which is a homolog of the eukaryotic SWI/SNF family, a set of
proteins whose members are involved in transcription activation,
50 nucleosome remodeling, and DNA repair. RapA forms a stable complex with RNAP,

5

56

10

15

20

25

30

35

40

45

50

55

as if it were a subunit of RNAP and it is possible that the ORF 8 product behaves similarly or in a dominant-negative fashion to inhibit the activity of RapA. Mutation of the essential *E. coli* dnaZX results in a block in DNA chain elongation during replication (Maki et al., 1988). The dnaZX gene has only one open reading frame for a 71-kDa polypeptide from which the two distinct DNA polymerase III holoenzyme subunits, tau (71 kDa) and gamma (47 kDa), are produced. The tau subunit is the precursor of the gamma subunit, and the gamma subunit is produced by a -1 frameshift causing early termination of translation (Tsuchihashi et al., 1990). These proteins show single-strand DNA binding properties that is ATPase (and dATPase) dependent and are thought to increase the processivity of the core DNA polymerase enzyme (Lee et al., 1987).

There are several Dp-1 ORFs which encode proteins predicted to play a role in cellular metabolic pathways. These include polypeptides involved in coenzyme PQQ synthesis (ORFs 20, 29, 38). Pyrrolo-quinoline quinone (PQQ) is the non-covalently bound prosthetic group of many quinoproteins catalysing reactions in the periplasm of Gram-negative bacteria. Most of these involve the oxidation of alcohols or aldose sugars. Interestingly, ORFs 20, 29, and 30 also show homology to the exoenzyme S regulon (Frank, 1997). Proteins encoded by the *P. aeruginosa* exoenzyme S regulon may be involved in a contact-mediated translocation mechanism to transfer anti-host factors directly into eukaryotic cells disrupting eukaryotic signal transduction through ADP-ribosylation (Frank, 1997).

There is also a protein with similarity to GTP cyclohydrolase I (ORF 21) and ORF 41 which shows homology to dUTPase (Table 31). GTP cyclohydrolase I is an enzyme that catalyzes the first reaction in the pathway for the biosynthesis of the pteridine, a cofactor of the monooxygenases of the aromatic amino acids. Disruption of the homologous gene in *Saccharomyces cerevisiae* leads to a recessive conditional lethality due to folinic acid auxotrophy, that can be complemented with the mammalian or bacterial GTP cyclohydrolase I enzymes (Nardese et al., 1996; Mancini et al., 1999).

ORF 16 shows high homology to autolysin. This region of the phage sequence was previously reported (Sheehan et al., 1997) and encompasses ~ 4 kbp of our sequence. The sequence published by (Sheehan et al., 1997) is shown in Table 32.

Thus, the present invention provides a nucleic acid sequence obtained from *Streptococcus* bacteriophage Dp-1 comprising at least a portion of a phage Dp-1 ORF, preferably an inhibitory ORF, and more preferably at least a portion of one of the genes described above with anti-microbial activity. For example, ORF 013 encodes a

protein with homology to the gamma subunit of DNA polymerase (dnaX gene). This protein may act in a dominant-negative fashion to sequester the host DNA polymerase for its own replication, thus inhibiting host DNA replication. The dnaX gene product is essential for *E. coli* replication (Kodaira et al., 1983).

In certain preferred embodiments of the present invention, the bacterial target of a bacteriophage inhibitor ORF product, e.g., an inhibitory protein or polypeptide, is encoded by a *Streptococcus* nucleic acid coding sequence from a host bacterium for bacteriophage Dp-1. As above, possible target sequences are described herein by reference to sequence source sites. The sequence encoding the target preferably corresponds to a *Streptococcus* nucleic acid sequence available from The Institute for Genomic Research (TIGR), or available from GenBank or other public database. The TIGR *Streptococcus* sequences are publicly available at The Institute for Genomics Research at URL: <http://www.tigr.org>

The amino acid sequence of a polypeptide target is readily provided by translating the corresponding coding region. For the sake of brevity, the sequences are not reproduced herein. Also, in preferred embodiments, a target sequence corresponds to a *Streptococcus pneumoniae* coding sequences corresponding to a sequence listed in Table 33 herein. Sequences for other Streptococcal species are also available from TIGR and/or from GenBank. The listing in Table 33 describes *Streptococcus* sequences currently deposited in GenBank. Again, for the sake of brevity, the sequences are described by reference to the GenBank entries instead of being written out in full herein. In cases where the TIGR or GenBank entry for a coding region is not complete, the complete sequence can be readily obtained by routine methods, e.g., by isolating a clone in a phage Dp-1 host *Streptococcus* sp. genomic library, and sequencing the clone insert to provide the relevant coding region. The boundaries of the coding region can be identified by conventional sequence analysis and/or by expression in a bacterium in which the endogenous copy of the coding region has been inactivated and using subcloning to identify the functional start and stop codons for the coding region.

In the various aspects of this invention involving Dp-1 sequences, preferably the sequence is preferably not contained in the sequence described in Sheehan et al., 1997 (Table 32).

Validating Identified Inhibitory Phage ORFs

A fifth step involves validating the identified phage inhibitor ORF by independent methods, and delineating further possible smaller segments of the ORFs

5

58

that have inhibitory activity. Several methods exist to validate the role of the identified ORF as an inhibitor ORF.

10

One example utilizes the creation of a mutant variant of the phage ORF in which the candidate ORF carries a partial or complete loss-of-function mutation that is measurable as compared with the non-mutant ORF. Comparison of the effects of expression of the loss of function mutant with the normal ORF provides confirmation of the identification of an inhibitor ORF where the loss-of-function mutant provides a measurably lower level of inhibition, preferably no inhibition. The loss of function may be conditional, *e.g.*, temperature sensitive.

15

Once validation of the inhibitor ORF is achieved, a bi-directional deletion analysis can be carried out using the same experimental system to identify the minimal polypeptide segment that has inhibitor activity. This may be carried out by a variety of means, *e.g.*, by exonuclease or PCR methodologies, and is used to determine if a relatively small segment of the ORF (*i.e.*, the product of the ORF) still possesses inhibitory activity when isolated away from its native sequence. If so, a portion of the ORF encoding this "active portion" can be used as a template for the synthesis of novel anti-microbial agents and further allowing derivation of the peptide sequence, *e.g.*, using modified peptides and/or peptidomimetics.

20

25

In creation of certain peptidomimetics, the peptide backbone is transformed into a carbon-based hydrophobic structure that can retain inhibitor activity against the bacterium. This is done by standard medicinal chemistry methods, typically monitored by measuring growth inhibition of the various molecules in liquid cultures or on solid medium. These mimetics can also represent lead compounds for the development of novel antibiotics.

30

35

Recently, a major effort has been undertaken by the pharmaceutical industry and their biotechnology partners for the sequencing of bacterial pathogen genomes. The rationale is that the systematic sequencing of the genome will identify all of the bacterial proteins and therefore this proteome will be the target for designing novel inhibitor antibiotics. Although systematic, this approach has several major problems. The first is that analysis of primary amino acid sequences of bacterial proteins does not immediately reveal which protein will be essential for viability of the bacterium, and target validation is thus a major issue. The second problem is one of redundancy, as several biochemical pathways are either structurally duplicated in bacteria (different isoforms of the same enzyme), or functionally duplicated by the presence of salvage pathways in the event of a metabolic block in one pathway (different nutritional conditions). The third is that even a valid target may not be structurally or

40

45

50

55

functionally amenable to inhibition by small molecules because of inaccessibility (sequestration of target).

Therefore, there is considerable interest within the pharmaceutical and biotechnology industry in identifying key targets for drug discovery amongst the mass of novel targets generated by large-scale genomic sequencing projects.

On the other hand, and underscoring the instant invention, the phages herein described have, over millions of years, evolved specific mechanisms to target such key biochemical pathways and proteins. In the few cases where inhibition by phages has been elucidated (e.g., see ref. 3), such bacterial targets are invariably rate-limiting in their respective biochemical pathways, are not redundant, and/or are readily accessible for inhibition by the phage (or by another inhibitory compound).

Therefore, the sixth step of this invention involves identifying the host biochemical pathways and proteins that are targeted by the phage inhibitory mechanisms.

Identifying, Validating, and Characterizing Bacterial Host Target Proteins and Affected Pathways

A rationale for this step is that the inhibitor ORF product from the phage physically interacts with and/or modifies certain microbial host components to block their function. Exemplary approaches which can be used to identify the host bacterial pathways and proteins that interact with, and preferably also are inhibited by, phage ORF product(s) are described below.

One approach is a genetic screen to determine physiological protein:protein interaction, for example, using a yeast two hybrid system. In this assay, the phage ORF is fused to the carboxyl terminus of the yeast Gal4 activation domain II (amino acids 768-881) to create a bait vector. A cDNA library of cloned *S. aureus* sequences which have been engineered into a plasmid where the *S. aureus* sequences are fused to the DNA binding domain of Gal4 is also generated. These plasmids are introduced alone, or in combination, into yeast strain Y190 - previously engineered with chromosomally integrated copies of the *E. coli lacZ* and the selectable *HIS3* genes, both under Gal4 regulation (Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). *Genes & Dev.* 7, 555-569). If the two proteins expressed in yeast interact, the resulting complex will activate transcription from promoters containing Gal4 binding sites. A *lacZ* and *His3* gene, each driven by a promoter containing Gal4 binding sites, have been integrated into the genome of the host yeast system used for measuring protein-protein interactions. Such a system provides a physiological environment in which to detect potential protein interactions. This system has been extensively used to identify novel protein-protein

5

60

10

15

20

25

30

35

40

45

50

55

interaction partners and to map the sites required for interaction (for example, to identify interacting partners of translation factors (Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). *Mol & Cell Biology* 18, 2697-2711), transcription factors (Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998). *Genes, Chromosomes & Cancer* 21, 217-222), and proteins involved in signal transduction (Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. *Nature*. 387, 921-924). This approach has also been used in many published reports to identify interaction between mammalian viral and mammalian cell proteins.

For example, the non-structural protein NS1 of parvovirus is essential for viral DNA amplification and gene expression and is also the major cytopathic effector of these viruses. A yeast two-hybrid screen with NS1 identified a novel cellular protein of unknown function that interacts with NS-1, called SGT, for small glutamine-rich tetratricopeptide repeat (TPR)-containing protein (Cziepluch C. Kordes E. Poirey R. Grewenig A. Rommelaere, J. and Jauniaux JC. (1998) *J Virol.* 72, 4149-4156). In another screen, the adenovirus E3 protein was recently shown to interact with a novel tumor necrosis factor alpha-inducible protein and to modulate some of the activities of E3 (Li Y. Kang J. and Horwitz M.S. (1998). *Mol & Cell Biol.* 18, 1601-1610). In yet another recent screen, the herpes simplex virus 1 alpha regulatory protein ICP0 was found to interact with (and stabilize) the cell cycle regulator cyclin D3 (Kawaguchi Y. Van Sant C. and Roizman B. (1997). *J Virol.* 71, 7328-7336).

Another two-hybrid system for identifying protein:protein interactions is commercially available from STRATEGENE™ as the CYTO-TRAP™ system (Chang et al., *Strategies Newsletter* 11(3), 65-68 (1998)(from Stratagene)). The system is a yeast-based method for detecting protein:protein interactions *in vivo*, using activation of the Ras signal transduction cascade by localizing a signal pathway component, human Sos (hSos), to its activation site in the yeast plasma membrane. The system uses a temperature-sensitive *Saccharomyces cerevisiae* mutant, strain cdc25H, which contains a point mutation at amino acid residue 1328 of the cdc25 gene. This gene encodes a guanyl nucleotide exchange factor which binds and activates Ras, leading to cell growth. The mutation in the cdc25 gene prevents host growth at 37°C, but at a permissive temperature of 25°C, growth is normal. The system utilizes the ability of (hSos) to complement the cdc25 defect and activate the yeast Ras signaling pathway. Once (hSos) is expressed and localized to the plasma membrane, the cdc25H yeast strain grows at 37°C. Localizing hSos to the plasma

membrane occurs through a protein:protein interaction. A protein of interest, or bait, is expressed as a fusion protein with hSos. The library, or target proteins are expressed with the myristylation membrane-localization signal. The yeast cells are then incubated under restrictive conditions (37°C). If the bait and the target protein interact, the hSos protein is recruited to the membrane, activating the Ras signaling pathway and allowing the cdc25H yeast strain to grow at the restrictive temperature.

The protein targets of phage inhibitory ORFs can also be identified using bacterial genetic screens. One approach involves the overexpression of a phage inhibitory protein in mutagenized bacterial host species, followed by plating the cells and searching for colonies that can survive the antimicrobial activity of the inhibitory ORF. These colonies are then grown, their DNA extracted, and cloned into an expression vector that contains a replicon of a different incompatibility group from the plasmid expressing the original ORF. This library is then introduced into a wild-type host bacterium in conjunction with an expression vector driving synthesis of the phage ORF, followed by selection for surviving bacteria. Thus, bacterial DNA fragments from the survivors presumably contain a DNA fragment from the original mutagenized host bacterial genome that can protect the cell from the antimicrobial activity of the inhibitory phage ORF. This fragment can be sequenced and compared with that of the bacterial host to determine in which gene the mutation lies. This approach enables one to determine the targets and pathways that are affected by the killing function.

A second approach is based on identifying protein:protein interactions between the phage ORF product and bacterial *S. aureus*, e.g., proteins using a biochemical approach based, for example, on affinity chromatography. This approach has been used, for example, to identify interactions between lambda phage proteins and proteins from their *E. coli* host (Sopta, M., Carthew, R.W., and Greenblatt, J. (1985) *J. Biol. Chem.* 260, 10353-10369). The phage ORF is fused to a peptide tag (e.g. glutathione-S-transferase ("GST"), 6xHIS, ("HIS") and/or calmodulin binding protein ("CPB")) within a commercially available plasmid vector that directs high level expression on induction of a suitably responsive promoter driving the fusion's expression. The translated fusion protein is expressed in *E. coli*, purified, and immobilized on a solid phase matrix via, for example the tag. Total cell extracts from the host bacterium, e.g., *S. aureus*, are then passed through the affinity matrix containing the immobilized phage ORF fusion protein; host proteins retained on the column are then eluted under different conditions of ionic strength, pH, detergents etc., and characterized by gel electrophoresis and other techniques. Appropriate controls are run to guard against nonspecific binding to the resin. Target proteins thus

recovered should be enriched for the phage protein/peptide of interest and are subsequently electrophoretically or otherwise separated, purified, sequenced, or biochemically analyzed. Usually sequencing entails individual digestion of the proteins to completion with a protease (e.g., trypsin), followed by molecular mass and amino acid composition and sequence determination using, for example, mass spectrometry, e.g., by MALDI-TOF technology (Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997). *Anal. Chem.* 69, 3995-4001).

The sequence of the individual peptides from a single protein are then analyzed by the bioinformatics approach described above to identify the *S. aureus* protein interacting with the phage ORF. This analysis is performed by a computer search of the *S. aureus* genome for an identified sequence. Alternatively, all tryptic peptide fragments of the *S. aureus* genome can be predicted by computer software, and the molecular mass of such fragments compared to the molecular mass of the peptides obtained from each interacting protein eluted from the affinity matrix. The responsible gene sequence can be obtained, for example by using synthetic degenerate nucleic acid sequences to pull out the corresponding homologous bacterial sequence. Alternatively, antibodies can be generated against the peptide and used to isolate nascent peptide/mRNA transcript complexes, from which the mRNA can be reverse transcribed, cloned, and further characterized using the procedures discussed herein.

A variety of other binding assay methods are known in the art and can be used to identify interactions between phage proteins and bacterial proteins or other bacterial cell components. Such methods that allow or provide identification of the bacterial component can be used in this invention for identifying putative targets.

Validation of the interaction between the phage ORF product and the bacterial proteins or other components can be obtained by a second independent assay (e.g., co-immunoprecipitation or protein-protein crosslinking experiments (Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). *Mol & Cell Biology* 18, 2697-2711; Brown, S. and Blumenthal, T. (1976). *Proc. Natl. Acad. Sci. USA* 73, 1131-1135)).

Finally, the essential nature of the identified bacterial proteins is preferably determined genetically by creating a constitutive or inducible partial or complete loss-of-function mutation in the gene encoding the identified interacting bacterial protein. This mutant is then tested for bacterial survival and replication.

The protein target of the phage inhibitor function can also be identified using a genetic approach. Two exemplary approaches will be delineated here. The first approach involves the overexpression of a predetermined phage inhibitor protein in mutagenized host bacteria, e.g., *S. aureus*, followed by plating the cells and searching

for colonies that can survive the inhibitor. These colonies will then be grown, their DNA extracted and cloned into an expression vector that contains a replicon of a different incompatibility group, and preferably having a different selectable marker than the plasmid expressing the phage inhibitor. Thus, host DNA fragments from the mutant that can protect the cell from phage ORF inhibition can be sequenced and compared with that of the bacterial host to determine in which gene the mutation lies. This approach allows rapid determination of the targets and pathways that are affected by the inhibitor.

Alternatively, the bacterial targets can be determined in the absence of selecting for mutations using an approach known as "multicopy suppression". In this approach, the DNA from the wild type host is cloned into an expression vector that can coexist, as previously described, with one containing a predetermined phage inhibitor. Those plasmids that contain host DNA fragments and genes that protect the host from the phage inhibitor can then be isolated and sequenced to identify putative targets and pathways in the host bacteria.

Regardless of the specific mode of identification, screening assays may additionally utilize gene fusions to specific "reporter genes" to identify a bacterial gene(s) whose expression is affected when the host target pathway is affected by the phage inhibitor. Such gene fusions can be used to search a number of small molecule compounds for inhibitors that may affect this pathway and thus cause cell inhibition. This approach will allow the screening of a large number of molecules on petri dishes or 96-well format by monitoring for a simple color change in the bacterial colonies. In this manner, we can validate host targets and classes of compounds for further study and clinical development. These inhibitors also represent lead compounds for the development of other antibiotics.

Bioinformatics and comparative genomics are preferably then applied to the identified bacterial gene products to predict biochemical function. The biochemical activity of the protein can be verified *in vitro* in cell free assays or *in vivo* in intact cells. *In vitro* biochemical assays utilizing cell-free extracts or purified protein are established as a basis for the screening and development of inhibitors.

These inhibitors, preferably small molecule inhibitors, may comprise peptides, antibodies, products from natural sources such as fungal or plant extracts or small molecule organic compounds. In general, small molecule organic compounds are preferred. These compounds may, for example, be identified within large compound libraries, including combinatorial libraries. For example, a plurality of compounds, preferably a large number of compounds can be screened to determine whether any of the compounds binds or otherwise disrupts or inhibits the identified bacterial target.

Compounds identified as having any of these activities can then be evaluated further in cell culture and/or animal model systems to determine the pharmacological properties of the compound, including the specific anti-microbial ability of the compound.

For mixtures of natural products, including crude preparations, once a preparation or fraction of a preparation is shown to have an anti-microbial activity, the active substance can be isolated and identified using techniques well known in the art, if the compound is not already available in a purified form.

Identified compounds possessing anti-microbial activity and similar compounds having structural similarity can be further evaluated and, if necessary, derivatized according to synthesis and/or modification methods available in the art selected as appropriate for the particular starting molecule.

Derivatization of identified anti-microbials

In cases where the identified anti-microbials above might represent peptidal compounds, the *in vivo* effectiveness of such compounds may be advantageously enhanced by chemical modification using the natural polypeptide as a starting point and incorporating changes that provide advantages for use, for example, increased stability to proteolytic degradation, reduced antigenicity, improved tissue penetration, and/or improved delivery characteristics.

In addition to active modifications and derivative creations, it can also be useful to provide inactive modifications or derivatives for use as negative controls or introduction of immunologic tolerance. For example, a biologically inactive derivative which has essentially the same epitopes as the corresponding natural antimicrobial can be used to induce immunological tolerance in a patient being treated. The induction of tolerance can then allow uninterrupted treatment with the active anti-microbial to continue for a significantly longer period of time.

Modified anti-microbial polypeptides and derivatives can be produced using a number of different types of modifications to the amino acid chain. Many such methods are known to those skilled in the art. The changes can include, for example, reduction of the size of the molecule, and/or the modification of the amino acid sequence of the molecule. In addition, a variety of different chemical modifications of the naturally occurring polypeptide can be used, either with or without modifications to the amino acid sequence or size of the molecule. Such chemical modifications can, for example, include the incorporation of modified or non-natural amino acids or non-amino acid moieties during synthesis of the peptide chain, or the post-synthesis modification of incorporated chain moieties.

The oligopeptides of this invention can be synthesized chemically or through an appropriate gene expression system. Synthetic peptides can include both naturally occurring amino acids and laboratory synthesized, modified amino acids.

Also provided herein are functional derivatives of anti-microbial proteins or polypeptides. By "functional derivative" is meant a "chemical derivative," "fragment," "variant," "chimera," or "hybrid" of the polypeptide or protein, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with a specific antibody, enzymatic activity or binding activity.

A "chemical derivative" of the complex contains additional chemical moieties not normally a part of the protein or peptide. Such moieties may improve the molecule's solubility, absorption, biological half-life, and the like. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, and the like. Moieties capable of mediating such effects are disclosed in Alfonso and Gennaro (1995). Procedures for coupling such moieties to a molecule are well known in the art. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, as described below.

Cysteiny residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteiny residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloro-mercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysiny and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysiny residues. Other suitable reagents for derivatizing primary amine-containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride;

trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine alpha-amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction carbodiimide ($R'-N-C-N-R'$) such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking component peptides to each other or the complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example; 1,1-bis (diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl) dithiolpropioimide yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E.,

Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex. Moieties capable of mediating such effects are disclosed, for example, in Alfonso and Gennaro (1995).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the protein or polypeptide having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide that either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a naturally occurring polypeptide by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

A functional derivative of a protein or polypeptide with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman et al., 1983, *DNA* 2:183; Sambrook et al., 1989) wherein nucleotides in the DNA coding sequence are modified such that a modified coding sequence is produced, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, components of functional derivatives of complexes with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art.

Insofar as other anti-microbial inhibitor compounds identified by the invention described herein may not be peptidal in nature, other chemical techniques exist to allow their suitable modification, as well, and according the desirable principles discussed above.

Administration and Pharmaceutical Compositions

For the therapeutic and prophylactic treatment of infection, the preferred method of preparation or administration of anti-microbial compounds will generally vary depending on the precise identity and nature of the anti-microbial being delivered. Thus, those skilled in the art will understand that administration methods known in the art will also be appropriate for the compounds of this invention.

The particularly desired anti-microbial can be administered to a patient either by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s). In treating an infection, a therapeutically effective amount of an agent or agents is administered. A therapeutically effective dose refers to that amount of the compound that results in amelioration of one or more symptoms of bacterial infection and/or a prolongation of patient survival or patient comfort.

Toxicity, therapeutic and prophylactic efficacy of anti-microbials can be determined by standard pharmaceutical procedures in cell cultures and/or experimental organisms such as animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50} . Compounds that exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized.

For any compound identified and used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. Such information can be used to more accurately determine useful doses in organisms such as plants and animals, preferably mammals, and most preferably humans. Levels in plasma may be measured, for example, by HPLC or other means appropriate for detection of the particular compound.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition (see *e.g.* Fingl et. al., in *The Pharmacological Basis of Therapeutics*, 1975, Ch. 1 p.1).

It should be noted that the attending physician would know how and when to terminate, interrupt, or adjust administration due to toxicity, organ dysfunction, or other systemic malady. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding

toxicity). The magnitude of an administered dose in the management of the disorder of interest will vary with the severity of the condition to be treated and the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and response of the individual patient. A program comparable to that discussed above also may be used in veterinary or phyto medicine.

Depending on the specific infection target being treated and the method selected, such agents may be formulated and administered systemically or locally, i.e., topically. Techniques for formulation and administration may be found in Alfonso and Gennaro (1995). Suitable routes may include, for example, oral, rectal, transdermal, vaginal, transmucosal, intestinal, parenteral, intramuscular, subcutaneous, or intramedullary injections, as well as intrathecal, intravenous, or intraperitoneal injections.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Use of pharmaceutically acceptable carriers to formulate identified anti-microbials of the present invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular those formulated as solutions, may be administered parenterally, such as by intravenous injection. Appropriate compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. Liposomes are spherical lipid bilayers with aqueous interiors. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external microenvironment and, because liposomes fuse with cell membranes, are efficiently

delivered into the cell cytoplasm. Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions, including those formulated for delayed release or only to be released when the pharmaceutical reaches the small or large intestine.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active anti-microbial compounds in water-soluble form.

Alternatively, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added.

The above methodologies may be employed either actively or prophylactically against an infection of interest.

Computer-related Aspects and Embodiments

In addition to the provision of compounds as chemical entities, nucleotide sequences, or fragments thereof at least 95%, preferably at least 97%, more preferably at least 99%, and most preferably at least 99.9% identical to phage inhibitor sequences can also be provided in a variety of additional media to facilitate various uses.

Thus, as used in this section, "provided" refers to an article of manufacture, rather than an actual nucleic acid molecule, which contains a nucleotide sequence of the present invention; e.g., a nucleotide sequence of an exemplary bacteriophage or a sequence encoding a bacterial target or a fragment thereof, preferably a nucleotide sequence at least 95%, more preferably at least 99% and most preferably at least 99.9% identical to such a bacteriophage or bacterial sequence, for example, to a polynucleotide of an unsequenced phage listed in Table 1, preferably of bacteriophage 77 (*S. aureus* host) or bacteriophage 3A (*S. aureus* host) or bacteriophage 96 (*S. aureus* host). Such an article provides a large portion of the particular bacteriophage genome or bacterial gene and parts thereof (e.g., a bacteriophage open reading frame (ORF)) in a form which allows a skilled artisan to examine and/or analyze the sequence using means not directly applicable to examining the actual genome or gene, or subset thereof as it exists in nature or in purified form as a chemical entity.

In one application of this aspect, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer

5

72

10

15

20

25

30

35

40

45

50

55

readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create an article of manufacture which includes one or more computer readable media having recorded thereon a nucleotide sequence or sequences of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can, for example, be presented in a word processing text file, formatted in commercially available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form a nucleotide sequence of an unsequenced bacteriophage, such as an exemplary bacteriophage listed in Table 1 or of a sequence encoding a bacterial target or a fragment thereof, preferably a nucleotide sequence at least 95%, more preferably at least 99% and most preferably at least 99.9% identical to such a bacteriophage or bacterial sequence, for example, to a polynucleotide of bacteriophage 77 (*S. aureus* host) or bacteriophage 3A (*S. aureus* host) bacteriophage

96 (*S. aureus* host), bacteriophage 44AHJD (*S. aureus* host), bacteriophage Dp-1 (*Streptococcus pneumoniae* host), or bacteriophage 182 (*Enterococcus* host) the present invention enables the skilled artisan to routinely access the provided sequence information for a wide variety of purposes.

Those skilled in the art understand that software can implement a variety of different search or analysis software which implement sequence search and analysis algorithms, e.g., the BLAST (Altschul et al., J. Mol. Biol. 215:403410 (1990) and BLAZE (Brutlag et al., Comp. Chem 17:203-207 (1993)) search algorithms. For example, such search algorithms can be implemented on a Sybase system and used to identify open reading frames (ORFs) within the bacteriophage genome which contain homology to ORFs or proteins from other viruses, e.g., other bacteriophage, and other organisms, e.g., the host bacterium. Among the ORFs discussed herein are protein encoding fragments of the bacteriophage genomes which encode bacteria-inhibiting proteins or fragments.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described. Such systems are designed to identify, among other things, useful fragments of the bacteriophage genomes.

As used herein, "a computer-based system" refers to the hardware, software, and data storage media used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input device, output device, and data storage medium or media. A skilled artisan will readily recognize that any of the currently available general purpose computer-based system are suitable for use in the present invention, as well as a variety of different specialized or dedicated computer-based systems.

As stated above, the computer-based systems of the present invention comprise data storage media having stored therein a nucleotide sequence of the present invention and the necessary hardware and software for supporting and implementing a search and/or analysis program.

As used herein, "data storage media" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search program" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means.

Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches and/or sequence analyses can be adapted for use in the present computer-based systems.

As used herein in connection with sequence searches and analyses, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Also, the target sequence length is preferably selected to include sequence corresponding to a biologically relevant portion of an encoded product, for example a region which is expected to be conserved across a range of source organisms. Preferably the sequence length of a target polypeptide sequence is from 5-100 amino acids, more preferably 7-50 or 7-100 amino acids, and still more preferably 10-80 or 10-100 amino acids. Preferably the sequence length of a target polynucleotide sequence is from 15-300 nucleotide residues, more preferably from 21-240 or 21-300, and still more preferably 30-150 or 30-300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length. Likewise, it may be desirable to search and/or analyze longer sequences.

As used herein, "a target-structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output devices can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output device ranks fragments of the bacteriophage or bacterial sequences possessing varying degrees of homology to the

5

75

10

15

20

25

30

35

40

45

50

55

target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing methods and/or devices and/or formats can be used to compare a target sequence or target motif with the sequence stored in data storage media to identify sequence fragments of the bacteriophage or bacterium in question. One skilled in the art can readily recognize that any one of the publicly available homology search programs can be used as the search program for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill, or later developed, also may be employed in this regard.

Figure 6 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well-known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

The data storage medium in which the sequence is embodied and the central processor need not be part of a single stand-alone computer, but may be separated so long as data transfer can occur. For example, the processor or processors being utilized for a search or analysis can be part of one general purpose computer, and the data storage medium can be part of a second general purpose computer connected to a network, or the data storage medium can be part of a network server. As another example the data storage medium can be part of a computer system or network accessible over telephone lines or other remote connection method.

EXAMPLES

Example 1. Growth of *Staph A* bacteriophage 77 and purification of genomic DNA.

The *Staphylococcus aureus* propagating strain (PS 77; ATCC #27699) was used as a host to propagate its respective phage 77 (ATCC # 27699-B1). Two rounds of plaque purification of phage 77 were performed on soft agar essentially as described in Sambrook et al (1989). Briefly, the PS 77 strain was grown overnight at 37°C in Nutrient broth [NB: 0.3% Bacto beef extract, 0.5% Bacto peptone (Difco Laboratories) and 0.5% NaCl (w/v)]. The culture was then diluted 20x in NB and incubated at 37°C until the $OD_{540} = .2$ (early log phase) with constant agitation. In order to obtain single plaques, phage 77 was subjected to 10-fold serial dilutions using phage buffer (1 mM $MgSO_4$, 5 mM $MgCl_2$, 80 mM NaCl and 0.1% Gelatin (w/v)) and 10 μ l of each dilution was used to infect 0.5 ml of the cell suspension in the presence of 400 μ g/ml $CaCl_2$. After incubation of 15 min at room temperature (RT), 2 ml of melted soft agar kept at 45°C (NB supplemented with 0.6% agar) was added to the mixture and poured onto the surface of 100 mm nutrient agar plates (0.3% Bacto Beef extract, 0.5% Bacto peptone, 0.5% NaCl and 1.5% Bacto agar (w/v)). After overnight incubation at 30°C, a single plaque was isolated, resuspended in 1 ml of phage buffer by end over end rotation for 2 hrs at 20°C, and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 30°C, a single plaque was isolated and used as a stock.

The propagation procedure for bacteriophage 77 was modified from the agar layer method of Swanstörn and Adams (1951). Briefly, the PS 77 strain was grown to stationary phase overnight at 37°C in Nutrient broth. The culture was then diluted twenty-fold in NB and incubated at 37°C until the $OD_{540} = .2$. The suspension (15×10^7 Bacteria) was then mixed with 15×10^5 plaque forming units (pfu) to give a ratio of 100-bacteria/phage particle in the presence of 400 μ g/ml of $CaCl_2$. After incubation for 15 min at 20°C, 7.5 ml of melted soft agar (NB plus 0.6% agar) were added to the mixture and poured onto the surface of 150 mm nutrient agar plates and incubated 16 hrs at 30°C. To collect the phage plate lysate, 20 ml of NB were added to each plate and the soft agar layer was collected by scrapping off with a clean microscope slide followed by shaking of the agar suspension for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 RPM (2,830xg) in a JA-10 rotor (Beckman) and the supernatant fluid (lysate) was collected and subjected to a treatment with 10 μ g/ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, the phage suspension was adjusted to 10% (w/v) PEG 8000 and

0.5 M of NaCl followed by incubation at 4°C for 16 hrs. The phage was recovered by centrifugation at 4,000 rpm (3,500xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman). The pellet was resuspended with 2 ml of phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin). The phage suspension was extracted with 1 volume of chloroform and further purified by centrifugation on a cesium chloride step gradient as described in Sambrook et al. (1989), using a TLS 55 rotor centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 h at 28,000 rpm (67,000xg) at 4°C. Banded phage was collected and ultracentrifuged again on an isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000xg) for 24 h at 4°C using a TLV rotor (Beckman). The phage was harvested and dialyzed for 4 h at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 mg/ml Proteinase K and 0.5% SDS and incubating for 1 h at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris pH 8.0, 1mM EDTA).

Example 2. DNA sequencing of Bacteriophage 77 genome

Four micrograms of phage 77 DNA was diluted in 200 µl of TE (10 mM Tris, [pH 8.0], 1 mM EDTA) in a 1.5 ml eppendorf tube and sonication was performed (550 Sonic Dismembrator™, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 s spaced by 15 s cooling in ice/water for 3 to 4 cycles. The sonicated DNA was then size fractionated by electrophoresis on 1% agarose gels utilizing TAE (1 x TAE is: 40 mM Tris-acetate, 1 mM EDTA [pH 8.0]) as the running buffer. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen), with a final elution of 50 µl of 1 mM Tris (pH 8.5).

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and the Klenow fragment of E. coli DNA polymerase I, as follows. Reactions were performed in a reaction mixture (final volume, 100 µl) containing sonicated phage DNA, 10 mM Tris-HCl [pH 8.0], 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 50 µg/ml BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of Klenow large fragment (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the

DNA was precipitated with ethanol and the final DNA pellet was resuspended in 20 μ l of H₂O.

Blunt-ended DNA fragments were cloned by ligation directly into the *Hinc* II site of pKSII+ vector (New England Biolabs) dephosphorylated by treatment with calf intestinal alkaline phosphatase (New England Biolabs)-treated pKS II+ vector (Stratagene). A typical ligation reaction contained 100 ng of vector DNA, 2 to 5 μ l of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 μ l containing 800 units of T4 DNA ligase (New England Biolabs) and was incubated overnight at 16°C. Transformation and selection of bacterial clones containing recombinant plasmids was performed in *E. coli* DH10 β according to standard procedures (Sambrook et al., 1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 μ l LB and 100 μ g/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers flanking the *Hinc* II cloning site of the pKS II+ vector. PCR amplification of foreign insert was performed in a 15 μ l reaction volume containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 μ M primer, 187.5 μ M each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 57°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using QIAprep™ spin miniprep kit (Qiagen).

The nucleotide sequence of the extremities of each recombinant clone was determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism Big Dye™ primer or ABI prism Big Dye™ terminator cycle sequencing ready reaction kit (Applied Biosystems). To ensure co-linearity of the sequence data and the genome, all regions of phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism Big Dye™ terminator cycle sequencing ready reaction kit.

Example 3. Bioinformatic management of primary nucleotide sequence from Phage 77.

Phage 77 sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of

the contigs. Phage DNA was used directly as sequencing template employing ABI prism BIG DYE™ terminator cycle sequencing ready reaction kit. The complete sequence of bacteriophage 77 is shown in Table 2.

A software program was developed and used on the assembled sequence of bacteriophage 77 to identify all putative ORFs larger than 33 codons. Other ORF identification software can also be utilized, preferably programs which allow alternative start codons. The software scans the primary nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG, GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI (<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c>) for the bacterial genetic code.

When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons (start and stop codons) is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence.

Sequence homology (BLAST) searches for each ORF are then carried out using an implementation of BLAST programs, although any of a variety of different sequence comparison and matching programs can be utilized as known to those skilled in the art. Downloaded public databases used for sequence analysis include:

- i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) *S. aureus* NCTC 8325 (<ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa>);
- vi) *streptococcus pyogenes* (<ftp://ftp.genome.ou.edu/pub/strep/strep-1k.fa>);
- vii) *Streptococcus pneumoniae* (ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- viii) *Mycobacterium tuberculosis* CSU#9 (ftp://ftp.tigr.org/pub/data/m_tuberculosis/TB_091097.Z) and
- ix) *pseudomonas aeruginosa* (<http://www.genome.washington.edu/pseudo/data.html>).

The results of the homology searches performed on the ORFs is shown in Table 5.

Example 4. Subcloning of Bacteriophage 77 ORFs into a Staph A inducible expression system.

The shuttle vector pT0021, in which the firefly luciferase (*lucFF*) expression is controlled by the *ars* (arsenite) promoter/operator (Tauriainen et al., 1997), was modified in the following fashion. Two oligonucleotides corresponding to a short antigenic peptide derived from the haemagglutinin protein of influenza virus (HA epitope tag) were synthesized (Field et al., 1988). The sense strand HA tag sequence (with *Bam*HI, *Sal*I and *Hind*III cloning sites) is:

5'-gatccgggtcgaccaagctTACCCATACGACGTCCTCCAGACTACGCCAGCTGA-3' (where upper case letters denote the nucleotide sequence of the HA tag); the antisense strand HA tag sequence (with a *Hind*III cloning site) is:

5'-agctTCAGCTGGCGTAGTCTGGGACGTCGTATGGGTAaagcttggtcgaccgg-3' (where upper case letters denote the sequence of the HA tag). The two HA tag oligonucleotides were annealed and ligated into pT0021 vector which had been digested with *Bam*HI and *Hind*III. This manipulation resulted in replacement of the *lucFF* gene by the HA tag. This modified shuttle vector containing the *arsenite* inducible promoter, the *arsR* gene, and HA tag was named pTHA. A diagram outlining our modification of pT0021 to generate pTHA is shown in Fig. 1A.

Each ORF, encoded by Bacteriophage 77, larger than 33 amino acids and having a Shine-Dalgarno sequence upstream of the initiation codon was selected for functional analysis for bacterial inhibition. In total, 98 ORFs were selected and screened as detailed below. A list of these is presented in Table 3. Each individual ORF, from initiation codon to last codon (excluding the stop codon), was amplified from phage genomic DNA using the polymerase chain reaction (PCR). For PCR amplification of ORFs, each sense strand primer targets the initiation codon and is preceded by a *Bam*HI restriction site (5'-cgggatcc-3') and each antisense oligonucleotide targets the penultimate codon (the one before the stop codon) of the ORF and is preceded by a *Sal*I restriction site (5'-gcgtcgaccg-3'). The PCR product of each ORF was gel purified and digested with *Bam*HI and *Sal*I. The digested PCR product was then gel purified using the Qiagen kit as described, ligated into *Bam*HI and *Sal*I digested pTHA vector, and used to transform *E. coli* bacterial strain DH10β (as described above). As a result of this manipulation, the HA tag is set inframe with the ORF and is positioned at the carboxy terminus of each ORF (pTHA/ORF clones). Recombinant pTHA/ORF clones were picked and their insert sizes were confirmed by PCR analysis

using primers flanking the cloning site. The names and sequences of the primers that were used for the PCR amplification were: HAF:

'TATTATCCAAAACCTTGAACA'; HAR: 'CGGTGGTATATCCAGTGATT'. The

sequence integrity of cloned ORFs was verified directly by DNA sequencing using primers HAF and HAR. In cases where verification of ORF sequence could not be achieved by one pass with the sequencing primers, additional internal primers were selected and used for sequencing.

Staphylococcus aureus strain RN4220 (Kreiwirth et al., 1983) was used as a recipient for the expression of recombinant plasmids. Electroporation was performed essentially as previously described (Schenk and Laddaga, 1992). Selection of recombinant clones was performed on Luria-Broth agar (LB-agar) plates containing 30 µg/ml of kanamycin.

For each ORF introduced in the pTHA plasmid, 3 independent transformants were isolated and used to individually inoculate cultures in 5 ml of TSB containing 30 µg/ml kanamycin, followed by growth to saturation (16 hrs at 30°C). An aliquot of this stationary phase culture was used to generate a frozen glycerol stock of the transformant (stored at -80°C). The remaining culture was used for plasmid DNA extraction. Bacterial cells were harvested by centrifugation at 3000 x g at 22°C for 5 min. The pellet was resuspended in 200 µl 25% sucrose containing 25U/ml of lysostaphin and incubated for 15 min at 37°C. Then, 400 µl of alkaline SDS solution (3% SDS, 0.2N NaOH) were added, well mixed and incubated for 7 min at room temperature. After the alkaline SDS treatment, 300 µl of ice-cold 3M sodium acetate pH 4.8 were added, and the mix is immediately spun at 13000g for 15 min at room temperature. The supernatant was transferred to a new 1.5 ml conical centrifuge tube and 650 µl of isopropanol (stored at room temperature) were added. The mix was then centrifuged at 13,000 x g for 5 min. The supernatant fluid was discarded, the pellet washed with 70% ethanol, and resuspended in 320 µl sterile distilled water.

The presence of individual phage 77 ORF DNA inserts in the plasmid was verified by PCR amplification using 1.5 µl transformant miniprep DNA in a PCR with primers flanking the cloning site of ORF in pTHA vector (HAF and HAR). The composition of the PCR reaction and the cycling parameters are identical to those employed for library screening described above.

Example 5. Functional assay for bacterial inhibitory activity of bacteriophage 77 ORFs.

The anti-microbial activity of individual phage 77 ORFs was monitored by two growth inhibitory assays, one on solid agar medium, the other in liquid medium.

In general, *Staphylococcus* bacteria transformed with expression plasmids containing individual ORFs were grown in normal TSA medium and stored in 19% glycerol. At pre-determined times, arsenite was added to the culture to induce transcription of the phage 77 ORFs cloned immediately downstream from an arsenite-inducible promoter in the pTHA expression plasmid.

The effect of ORF induction on bacterial growth characteristics was then monitored and quantitated. The growth inhibition assay on solid medium was performed by streaking pTHA/ORF containing *S. aureus* transformant onto LB-Kn and TSA-Kn plates containing increasing concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M). Arsenite is used to induce the expression of cloned DNA in pTHA vector. In parallel, 3 μ l of 1/10 and 1/100 dilutions of the frozen cultures of the pTHA/ORF transformants were spotted as single drops onto LB-Kn and TSA-Kn plates containing increasing concentration of sodium arsenite (0; 2.5; 5; and 7.5 μ M). The plates were then incubated 16 hrs at 37°C, and the effect of arsenite-induced ORF expression on bacterial growth was monitored and quantitated by comparing the extent to that seen in control plates. As positive controls for growth inhibition, the *holin/lys* genes of the *Staphylococcus aureus* phage Twort (Loessner et al., 1998) was subcloned into the pTHA *ars* inducible vector and used.

For the growth inhibition assay in liquid medium, stationary phase cultures were prepared by inoculating 2.5ml TSB-Kn with frozen *S. aureus* RN4220 transformants containing phage 77 ORFs cloned in pTHA vector followed by incubation for 16 hrs at 37°C. These cultures were then diluted 1/100 in the same medium, and the bacteria were allowed to grow for 2 hrs at 37°C to reach early log phase. 150 μ l of such culture were then mixed with 2.35 ml TSB-Kn medium with or without arsenite (the final concentration of arsenite in the medium was 0 or 5 μ M arsenite). After 3.5 hrs incubation at 37°C with shaking at 250 rpm, 100 μ l of bacterial culture was removed from each tube for OD₆₆₀ measurement. Serial ten-fold dilutions of the culture in buffered saline solution (0.85% NaCl) were then spotted onto TSB-Kn plates. The plates were incubated at 37°C 16 hrs and the number of surviving colonies counted the following day. The growth inhibitory property of individual ORFs was then quantitated by comparing CFU numbers under normal or arsenite-induction conditions. A schematic flow of the inhibition analysis is shown in Fig. 3 (also applicable to inhibition analysis for the other phage and bacteria pointed out herein). Inhibition results are shown in Figures 4A-C.

Example 6: Identification of Cecropin Signature Motif in *Staphylococcus aureus* Bacteriophage 3A ORF

The genome for *S. aureus* bacteriophage 3A was determined and the sequence was analyzed essentially as described for bacteriophage 77 in the examples above. Upon blast analysis of the identified open reading frames of phage 3A, the presence of an amino acid sequence corresponding to a cecropin signature motif was observed.

- 5 This motif (WDGHKTLEK) is located at position aa 481-489. Cecropins were originally identified in proteins from the cecropia moth and are recognized as potent antibacterial proteins that constitute an important part of the cell-free immunity of insects. Cecropins are small proteins (31-39 amino acid residues) that are active against both Gram-positive and Gram-negative bacteria by disrupting the bacterial membranes. Although the mechanisms by which the cecropins cause cell death are not fully understood, it is generally thought to involve channel formation and membrane destabilization.

- 10 The identification of a motif corresponding to a known inhibitor suggests that the product of ORF002 is also an inhibitory compound. Such inhibitory activity can be confirmed as described herein or by other methods known in the art. Confirmation of the inhibitory activity would indicate that the ORF product could serve as the basis for construction of mimetic compounds and other inhibitors directed to the target of the ORF002 product.

- 15 Boman & Hultmark, 1987, *Ann. Rev. Microbiol.* 41:103-126.
20 Boman, 1991, *Cell* 65:205-207.
Boman et al., 1991, *Eur. J. Biochem.* 201:23-31.
Wang et al., *J. Biol. Chem.* 273:27438-27448.

Example 7. Growth of *Staphylococcus aureus* bacteriophage 44AHJD:

- 25 *Staphylococcus aureus* propagating strain (PS 44A) (Felix d'Herelle Reference Centre #HER 1101) was used as a host to propagate its respective phage 44AHJD (Felix d'Herelle Reference Centre #HER 101). Two rounds of plaque purification of phage 44AHJD were performed on soft agar essentially as described in Sambrook *et al.* (1989). Briefly, the *Staphylococcus aureus* PS strain was grown overnight at 37°C in Nutrient Broth [NB: 3 g Bacto Beef Extract, 5 g Bacto peptone per liter, (Difco Laboratories # 0003-17-8), supplemented with 0.5% NaCl]. The culture was then diluted 20 fold in NB and incubated at 37°C until an OD₅₄₀ of 0.2. In order to obtain single plaques, phage 44AHJD was subjected to 10-fold serial dilutions using the phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin) and 10 µl were used to infect 0.5 ml of the cell suspension in the presence of 400 µg/ml of

CaCl₂. After incubation of 15 min at room temperature, 2 ml of melted soft agar (NB supplemented with 0.6% of agar) were added to the mixture and poured onto the surface of 100 mm nutrient agar plates (3 g Bacto Beef extract, 5 g Bactopeptone, 0.5% NaCl and 15 g of Bacto agar per liter (Difco Laboratories # 0001-17-0). After overnight incubation at 37°C, a single plaque was isolated, resuspended in 1 ml of phage buffer by end over end rotation for 2 h at room temperature and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 37°C, a single plaque was isolated and used as a stock.

Large scale purification of bacteriophage and preparation of phage DNA was as follows.

The propagation method was carried out by using the agar layer method described by Swanstörn and Adams (1951). Briefly, the PS 44A strain was grown to stationary phase overnight at 37°C in Nutrient Broth. The culture was then diluted 20x in NB and incubated at 37°C until the A₅₄₀ = 0.2. The suspension (15x10⁷ Bacteria) was then mixed with 15x10⁷ phage particles to give a ratio of 100-bacteria/phage particle in the presence of 400 µg/ml of CaCl₂. After incubation of 15 min at room temperature, 7.5 ml of melted soft agar were added to the mixture and poured onto the surface of 150 mm nutrient agar plates and incubated overnight at 37°C. To collect the lysate, 20 ml of NB were added to each plate and the soft agar layer was collected by scrapping off with a clean microscope slide and shaken vigorously for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 rpm (2,830 xg) using a JA-10 rotor (Beckman) and the supernatant (lysate) is collected and subjected to a treatment with 10 µg/ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, 10% (w/v) of PEG 8000 and 0.5 M of NaCl were added to the lysate and the mixture was incubated on ice for 16 h. The phage was recovered by centrifugation at 4,000 rpm (3,500 xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman).

The pellet was resuspended with 2 ml of phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin). The phage suspension was extracted with 1 volume of chloroform and further purified by centrifugation on a preformed cesium chloride step gradient as described in Sambrook *et al.* (1989), using a TLS 55 rotor and centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 h at 28,000 rpm (67,000 xg) at 4°C. Banded phage was collected and ultracentrifuged again on an

isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000 x g) for 24 h at 4°C using a TLV rotor (Beckman). The phage was harvested and dialyzed for 4 h at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 µg/ml Proteinase K and 0.5% SDS and incubating for 1 h at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris-HCl [pH 8.0], 1mM EDTA).

Example 8. DNA sequencing of the Bacteriophage 44 AHJD genome.

Four mg of phage DNA was diluted in 200 µl of TE pH 8.0 in a 1.5 ml eppendorf tube and sonication was performed (550 Sonic Dismembrator, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 s spaced by 15 s cooling in ice/water for 3 to 4 cycles and size fractionated on 1% agarose gels. The sonicated DNA was then size fractionated by gel electrophoresis. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen) and eluted in 50 µl of 1mM Tris-HCl [pH 8.5].

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and the Klenow fragment of *E. coli* DNA polymerase I as follows. Reactions were performed in a final volume of 100 µl containing DNA, 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 5 µg BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of Klenow fragment (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the DNA was ethanol precipitated and resuspended in 20 µl of H₂O.

Cloning of the sonicated phage DNA into pKSII vector and transformation:

Blunt-ended DNA fragments were cloned by ligation directly into the *HincII* site of the pKSII vector (Stratagene) dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs). A typical reaction contained 100 ng of vector, 2

to 5 µl of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 µl containing 800 units of T4 DNA ligase (New England Biolabs) overnight at 16°C. Transformation and selection of positive clones was performed in the host strain DH10 β of *E. coli* using ampicillin as a selective antibiotic as described in Sambrook *et al.* (1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 µl LB and 100 µg/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers flanking the *HincII* cloning site of the pKS vector. PCR amplification of the potential foreign inserts was performed in a 15 µl reaction volume containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 mM primer, 187.5 µM each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 58°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using the QIAprep™ spin miniprep kit (Qiagen).

The nucleotide sequence of the extremities of each recombinant clone was determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism BigDye™ primer cycle sequencing (21M13 primer: #403055)(M13REV primer: #403056) or ABI prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). To ensure co-linearity of the sequence data and the genome, all regions of the phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism BigDye™ terminator cycle sequencing ready reaction kit.

Example 9. Bioinformatic management of primary nucleotide sequence.

Sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of the contigs. Phage DNA was used directly as sequencing template employing ABI

prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). The complete sequence of *Staphylococcus aureus* bacteriophage 44AHJD is shown in Table 16.

A software program was used on the assembled sequence of bacteriophage 44AHJD to identify all putative ORFs larger than 33 codons. The software scans the primary nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG, GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI (<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c>) for the bacterial genetic code. When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence. The predicted ORFs for bacteriophage 44AHJD are listed in Tables 17 & 18.

Sequence homology searches for each ORF were carried out using an implementation of blast programs. Downloaded public databases used for sequence analysis include:

- (i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- 25 iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- 40 iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) *Staphylococcus aureus* NCTC 8325 (<ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa>);
- 45 vi) *Staphylococcus pyogenes* (ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- 30 vii) PRODOM (ftp://ftp.toulouse.inra.fr/pub/prodom/current_release/prodom99_1.förblast.gz);
- 50 viii) DOMO (<ftp://ftp.infobiogen.fr/pub/db/domo/>);

ix) TREMBL ([ftp://www.expasy.ch/databases/sp_tr_nrd/fasta/](http://www.expasy.ch/databases/sp_tr_nrd/fasta/))

The results of the homology searches performed on the ORFs of bacteriophage 44AHJD are shown in Tables 19 & 20.

5 Example 10. Sub-Cloning of Bacteriophage 44 AHJD ORFs.

Expression preferably utilizes a shuttle expression vector which is arranged such that expression of the exogenous bacteriophage 44 AHJD ORF sequence is inducible. For example, the shuttle vector pT0021, in which the firefly luciferase (*lucFF*) expression is controlled by the *ars* (arsenite) promoter/operator (Tauriainen et al., 1997), can be modified in the following fashion. Two oligonucleotides corresponding to a short antigenic peptide derived from the hemagglutinin protein of influenza virus (HA epitope tag) were synthesized (Field et al., 1988). The sense strand HA tag sequence (with *Bam*HI, *Sal*I and *Hind*III cloning sites) is:
 5'-gatcccggtcgaccaagcttTACCCATACGACGTCCAGACTACGCCAGCTGA-3'
 (where upper case letters denote the nucleotide sequence of the HA tag); the antisense strand HA tag sequence (with a *Hind*III cloning site) is:
 5'-agctTCAGCTGGCGTAGTCTGGGACGTCGTATGGGTAaagcttggtcgaccgg-3'
 (where upper case letters denote the sequence of the HA tag). The two HA tag oligonucleotides were annealed and ligated into pT0021 vector which had been digested with *Bam*HI and *Hind*III. This manipulation resulted in replacement of the *lucFF* gene by the HA tag. This modified shuttle vector containing the *arsenite* inducible promoter, the *arsR* gene, and HA tag was named pTHA. A diagram outlining our modification of pT0021 to generate pTHA is shown in Fig. 1A (another useful vector construct is shown in Fig. 1B).

Each ORF, encoded by Bacteriophage 44 AHJD, larger than 33 amino acids and having a Shine-Dalgarno sequence upstream of the initiation codon can be selected for functional analysis for bacterial inhibition. Each individual ORF, from initiation codon to last codon (excluding the stop codon), can be amplified from phage genomic DNA using the polymerase chain reaction (PCR). For PCR amplification of ORFs, each sense strand primer targets the initiation codon and is preceded by a *Bam*HI restriction site ("cgggatcc") and each antisense oligonucleotide targets the pentultimate codon (the one before the stop codon) of the ORF and is preceded by a *Sal*I restriction site ("gcgtcgaccg"). The PCR product of each ORF can be gel

5 purified and digested with *Bam*HI and *Sa*II. The digested PCR product can then be
gel purified using the Qiagen kit as described, ligated into *Bam*HI and *Sa*II digested
10 pTHA vector, and used to transform *E. coli* bacterial strain DH10 β (as described
above). As a result of this manipulation, the HA tag is set inframe with the ORF and is
5 positioned at the carboxy terminus of each ORF (pTHA/ORF clones). Recombinant
pTHA/ORF clones will be picked and their insert sizes were confirmed by PCR
15 analysis using primers flanking the cloning site. The following primers can be used
for PCR amplification: HAF: 'TATTATCCAAAACCTGAACA'; HAR:
'CGGTGGTATATCCAGTGATT'. The sequence integrity of cloned ORFs can be
10 verified directly by DNA sequencing using primers HAF and HAR. In cases where
20 verification of ORF sequence can not be achieved by one pass with the sequencing
primers, additional internal primers will be selected and used for sequencing.

Staphylococcus aureus strain RN4220 (Kreiwirth et al., 1983) will be used as
25 a recipient for the expression of recombinant plasmids. Electoporation will be
15 performed essentially as previously described (Schenk and Laddaga, 1992). Selection
of recombinant clones will be performed on Luria-Broth agar (LB-agar) plates
containing 30 μ g/ml of kanamycin.

30 Alternatively, a constitutive promoter can be used to drive expression of the
introduced ORF, and compare cell growth to control bacterial cells containing the
20 parental vector lacking any introduced phage ORF. Recombinant plasmids will be
introduced into *Staphylococcus aureus* strain RN4220 (Kreiwirth et al., 1983) using
35 electoporation as previously described (Schenk and Laddaga, 1992).

Cloning of ORFs with a Shine-Dalgarno sequence

30 ORFs with a Shine-Dalgarno sequence are selected for functional analysis of
25 bacterial killing. Each ORF, from initiation codon to last codon (excluding the stop
40 codon), can be amplified by PCR from phage genomic DNA. For PCR amplification
of ORFs, each sense strand primer starts at the initiation codon and is preceded by a
restriction site and each antisense strand starts at the last codon (excluding the stop
45 codon) and is preceded by a different restriction site. The PCR product of each ORF
30 will be gel purified and digested with the restriction enzymes with sites contained on
the PCR oligonucleotides. The digested PCR product is then gel purified using the
Qiagen kit, ligated into the modified shuttle vector, and used to transform bacterial
50 strain DH10. Recombinant clones are then picked and their insert sizes confirmed by

PCR analysis using primers flanking the cloning site as well as restriction digestion. The sequence fidelity of cloned ORFs can be verified by DNA sequencing using the same primers as used for PCR. In the cases that the verification of ORFs can not be achieved by one path of sequencing using primers flanking the cloning site internal primers can be selected and used for sequencing. Recombinant plasmids can be introduced into *Staphylococcus aureus* strain RN4220 (Kreiwirth et al., 1983) using electroporation as previously described (Schenk and Laddaga, 1992).

Induction of gene expression from the *ars* promoter.

If an inducible promoter is used, e.g., the *ars* promoter, induction can be assessed, for example, in either of the two methods.

1. Screening on agar plates

The functional identification of killer ORFs can be performed by spreading an aliquot of *S. aureus* transformed cells containing phage 44 AHJD ORFs onto agar plates containing different concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M). The plates are incubated overnight at 37°C, after which a growth inhibition of the ORF transformants on plates that contain arsenite are compared to plates without arsenite.

2. Quantification of growth inhibition in liquid medium

Cells containing different recombinant plasmids can be grown for overnight at 37°C in LB medium supplemented with the appropriate antibiotic selection. These are then diluted to the mid log phase ($OD_{540}=2$) with fresh media containing antibiotic and transferred to 96-well microtitration plates (100 μ l/well). Inducer is then added at different final concentrations (ranging from 2.5 to 10 μ M) and the culture incubated for an additional 2 hrs at 37°C. The effect of expression of the phage 44 AHJD ORFs on bacterial cell growth is then monitored by measuring the OD_{540} and comparing the rate of growth to the culture not containing inducer. [As positive controls for growth inhibition, the *kilA* gene of phage lambda (Reisinger, GR., Rietsch, A., Lubitz, W. and Blasi, U. 1993 *Virology* #193: 1033-1036), and the *holin/lysins* genes of the *Staphylococcus aureus* phage Twort (Loessner, MJ., Gaeng, S., Wendlinger, G., Maier, SK. and Scherer, S. 1998. *FEMS Microbiology Letters* #162:265-274) can be subcloned into the *ars* inducible vector. An aliquot of the induced and uninduced culture can also be plated out on agar plates containing an appropriate antibiotic-selection but lacking inducer. Following incubation overnight at 37°C, the number of

colonies is counted. Any ORF showing bacteriostatic activity will show a lower, but detectable, number of colonies on the agar plates when grown in the presence of inducer as compared to when grown in the absence of inducer. Any ORF showing full bacteriocidal activity will show no colonies on the agar plates, when grown in the presence of inducer as compared to when grown in the absence of inducer.

REFERENCES

Ackermann, H-W. and DuBow, M. S. (1987). Viruses of Prokaryotes. Volumes I and II. CRC Press, Boca Raton, Florida.

Tenover, F.C. and McGowan Jr., J.E. (1998). Bacterial Infections of Humans. Epidemiology and Control. (A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 83-93.

Rusterholtz, K., and Pohlschroder, M. (1999). Cell 96, 469-470.

Gray, B.M. (1998). Bacterial Infections of Humans. Epidemiology and Control. (A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 673-711.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press.

Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J.

Rost B.I and Sander C. (1996). Ann. Rev. Biophys. Biomol. Struct. 25, 113-136.

Martin, A.C., Lopez, R., Garcia, P. (1998). J Bacteriol 180, 210-217.

Steiner, M., Lubitz, W., Blasi, U. (1993). J. Bacteriol. 175, 1038-1042.

Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). Genes Dev. 7, 555-569.

Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). Mol Cell Biol. 18, 2697-2711.

5

10

Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998). *Genes, Chromosomes & Cancer* 21, 217-222.

15

- 5 Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. (1997). *Nature* 387, 921-924.

20

- 10 Karimova, G., Pidoux, J., Ullmann, A., Ladant, D. (1998) *Proc. Natl. Acad. Sci.* 95, 5752-5756.

Sopta, M., Carthew, R.W., and Greenblatt, J. (1995) *J. Biol. Chem.* 260, 10353-10369.

15

Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997). *Anal. Chem.* 69, 3995-4001.

25

- 20 Swanström, M. and Adams, M.H. (1951). *Proc. Soc. Exptl. Biol. & Med.* 78: 372-375.

30

Røder, B.L., Wandall, D. A., Frimødt-Møller, N., Epersen, F., Skinhøj, P. and Rosdahl, T. (1999). *Arch. Intern. Med.* 159: 462-469.

- 25 Sanabria, T.J., Albert, J.S., Goldberg, R., Pape, L.A. and Cheeseman, S.H. (1990). *Arch. Intern. Med.* 150: 1305-1309.

35

Frimødt-Møller, N., Epersen, F., Skinhøj, P. and Rosdahl, V.T. (1997). *Clin. Microbiol. Infect.* 3: 297-305.

30

Harbath, S., Rutschmann, O., Sudre, P. and Pittet, D. (1998). *Arch. Intern. Med.* 158: 182-189.

40

- 35 Steinberg, J.P., Clark, C.C. and Hackman, B.O. (1996). *Clin. Infect. Dis.* 23: 255-259.

45

Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lemer, R.A., and Wigler, M. (1988). Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* 8: 2159-2165.

40

Kreiswirth, BN., Lofdahl, S., Belley, MJ., O'Reilly, M., Shlievert, PM., Bergdoll, MS. and Novicks, RP. (1983) *Nature* 305: 709-712.

50

55

Schenk, S. and Laddaga, RA. (1992) FEMS Microbiology Letters 94: 133-138.

Cohen, M.L. (1992) Science 257, 1050-1055.

Example 11. Growth of *Enterococcus* bacteriophage 182 and purification of genomic DNA.

The *Enterococcus* propagating strain (PS) (*Enterococcus* sp. Group D, Felix d'Herelle Reference Centre #HER 1080) was used as host to propagate its respective phage 182 (Felix d'Herelle Reference Centre #HER 80). Two rounds of plaque purification of phage 182 were performed on soft agar essentially as described in Sambrook *et al.* (1989). Briefly, the *Enterococcus* sp. PS strain was grown overnight at 37°C in Tryptic Soy Broth [TSB: 17 g Bacto tryptone, 3 g Bacto soytone, 2.5 g Bacto dextrose, 5 g Sodium chloride, and 2.5 g Dipotassium phosphate per liter (Difco Laboratories (#0370-17-3))]. The culture was then diluted 20 fold in TSB and incubated at 37°C until the $OD_{540} = 0.2$ (early log phase) with constant agitation. In order to obtain single plaques, phage 182 was subjected to 10 fold serial dilutions using the phage buffer (1 mM $MgSO_4$, 5 mM $MgCl_2$, 80 mM NaCl and 0.1% Gelatin (w/v)) and 10 l of each dilution was used to infect 0.5 ml of the bacterial cell suspension. After incubation at 15 min at 37°C, 2 ml of melted soft agar (TSB supplemented with 0.6% agar) was added to the mixture and poured onto the surface of 100 mm Tryptic Soy Agar plates [TSA: 15 g Tryptone peptone, 5 g Soytone peptone, 5 g Sodium chloride and 15 g of Agar per liter (Difco Laboratories #0369-17)]. After overnight incubation at 37°C, a single plaque was isolated, resuspended in 1 ml of phage buffer by end over end rotation for 2 hrs at room temperature, and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 37°C, a single plaque was isolated and used as a stock for all subsequent manipulations.

The propagation procedure for bacteriophage 182 was modified from the agar layer method of Swanstörn and Adams (1951). Briefly, the *Enterococcus* sp. PS strain was grown to stationary phase overnight at 37°C in TSB. The culture was then diluted 20 fold in TSB and incubated at 37°C until the $A_{540} = 0.2$. The suspension (15×10^7 Bacteria) was then mixed with 15×10^4 plaque forming units (pfu) to give a

ratio of 100-bacteria/pfu. After incubation of 15 min at 37°C, 7.5 ml of melted soft agar (TSB plus 0.6% agar) were added to the mixture and poured onto the surface of 150 mm TSA plates and incubated 16 hrs at 37°C. To collect the plate lysate, 20 ml of TSB were added to each plate and the soft agar layer was collected by scrapping off with a clean microscope slide followed by vigorous shaking of the agar suspension for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 rpm (2,830 xg) using a JA-10 rotor (Beckman) and the supernatant fluid (lysate) is collected and subjected to a treatment with 10 µg/ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, the phage suspension was adjusted to 10% (w/v) of PEG 8000 and 0.5 M of NaCl followed by incubation at 4°C for 16 hrs. The phage was recovered by centrifugation at 4,000 rpm (3,500 xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman). The pellet was resuspended with 2 ml of phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin). The phage suspension was extracted with 1 volume of chloroform and further purified by centrifugation on a cesium chloride step gradient as described in Sambrook *et al.* (1989), using a TLS 55 rotor and centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 hrs at 28,000 rpm (67,000 xg) at 4°C. Banded phage was collected and ultracentrifuged again on an isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000 xg) for 24 hrs at 4°C using a TLV rotor (Beckman). The phages were harvested and dialyzed for 4 hrs at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 g/ml Proteinase K and 0.5% SDS and incubating for 1 hr at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris-HCl [pH 8.0], 1mM EDTA).

Example 12. DNA sequencing of the Bacteriophage 182 genome.

Four micrograms of phage DNA was diluted in 200 µl of TE (10 mM Tris, [pH 8.0], 1 mM EDTA) in a 1.5 ml eppendorf tube and sonication was performed (550 Sonic Dismembrator, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 s spaced by 15 s cooling in ice/water for 3 to 4

cycles. The sonicated DNA was then size fractionated by electrophoresis on 1% agarose gels utilizing TAE (1 x TAE is: 40 mM Tris-acetate, 1 mM EDTA [pH 8.0]) as the running buffer. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen), with a final elution of 50 µl of 1 mM Tris [pH 8.5].

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and the Klenow fragment of *E. coli* DNA polymerase I, as follows. Reactions were performed in a reaction mixture (final volume, 100 µl) containing sonicated phage DNA, 10 mM Tris-HCl [pH 8.0], 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 50 µg/ml BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of the Klenow large fragment of DNA polymerase I (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the DNA was precipitated with ethanol and the final DNA pellet resuspended in 20 µl of H₂O.

Blunt-ended DNA fragments were cloned by ligation directly into the *Hinc* II site of the pKSII+ vector (New England Biolabs) dephosphorylated by treatment with calf intestinal alkaline phosphatase (New England Biolabs). A typical ligation reaction contained 100 ng of vector DNA, 2 to 5 µl of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 µl containing 800 units of T4 DNA ligase (New England Biolabs) and was incubated overnight at 16°C. Transformation and selection of bacterial clones containing recombinant plasmids was performed in *E. coli* DH10β according to standard procedures (Sambrook *et al.*, 1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 µl LB and 100 µg/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers flanking the *Hinc* II cloning site of the pKS vector. PCR amplification of the potential foreign inserts was performed in a 15 µl reaction volume containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 µM primer, 187.5 µM each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec

denaturation at 94°C, 30 sec annealing at 58°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using the QIAprep™ spin miniprep kit (Qiagen).

The nucleotide sequence of the extremities of each recombinant clone was determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism Big Dye™ primer cycle sequencing (21M13 primer: #403055)(M13REV primer: #403056) or ABI prism Big Dye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). To ensure co-linearity of the sequence data and the genome, all regions of the phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism BigDye™ terminator cycle sequencing ready reaction kit.

Example 13. Bioinformatic management of primary nucleotide sequence.

Sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of the contigs. Phage DNA was used directly as sequencing template employing ABI prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). The complete sequence of *Enterococcus* bacteriophage 182 is shown in Table 21.

A software program was used on the assembled sequence of bacteriophage 182 to identify all putative ORFs larger than 33 codons. The software scans the primary nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG, GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI(<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintec?mode=c>) for the bacterial genetic code. When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the

next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence. The predicted ORFs for bacteriophage 182 are listed in Tables 22 & 23. Sequence homology searches for each ORF were carried out using an implementation of BLAST programs. Downloaded public databases used for sequence analysis include:

- (i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) staphylococcus aureus NCTC 8325 (<ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa>);
- vi) streptococcus pyrogenes (ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- vii) PRODOM (ftp://ftp.toulouse.inra.fr/pub/prodom/current_release/prodom99.1.forblast.gz);
- viii) DOMO (<ftp://ftp.infobiogen.fr/pub/db/domo/>);
- ix) TREMBL (ftp://www.expasy.ch/databases/sp_tr_nrd/fasta/)

The results of the homology searches performed on the ORFs of bacteriophage 182 are shown in Tables 24 & 26.

Example 14. Sub-Cloning of Bacteriophage 182 ORFs.

Preparation of the shuttle expression vector

Expression preferably utilizes a shuttle expression vector which is arranged such that expression of the exogenous bacteriophage 182 ORF sequence is inducible. For example, the plasmid pND50 replicates in *E. coli*, *E. faecalis*, and *S. aureus* (Yamagishi, J., Kojima, T., Oyamada, Y., Fujimoto, K., Hattori, H., Nakamura, S., and Inoue, M. 1996. *Antimicrob. Agents Chemother.* 40, 1157-1163). This plasmid can be modified by conventional techniques to insert the inducible arsenite promoter, derived from the shuttle vector pT0021, in which the firefly luciferase (*lucFF*)

expression is controlled by the *ars* promoter/operator from a *S. aureus* plasmid (Tauriainen, S., Karp, M., Chang, W and Virta, M. (1997). Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite. *Appl. Environ. Microbiol.* 63:4456-4461). This modified shuttle vector will contain the *ars* promoter, *arsR* gene and a cloning site for introduction of individual phage ORFs downstream from a shine-dalgarno sequence.

Other inducible regulatory sequences can be utilized instead of the arsenite-inducible system. An example is a nisin-inducible system. The *nisA* promoter activity is dependent on the proteins NisR and NisK, which constitute a two-component signal transduction system that responds to the extracellular inducer nisin. The nisin sensitivity and inducer concentration required for maximal induction varies among the strains, but is functional in *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Enterococcus faecalis*, and *Bacillus subtilis*. Significant induction of the *nisA* promoter (10- to 60-fold induction) can be obtained in all of the species. A vector containing this promoter was published as Eichenbaum Z, Federle MJ, Marra D, de Vos WM, Kuipers OP, Kleerebezem M, and Scott JR (1998) *Appl Environ Microbiol* 64, 2763-2769. Other vectors, e.g., plasmids, can also be utilized which will allow replication and transcription in *Enterococcus*.

Alternatively, a constitutive promoter can be used (e.g., the β -lactamase promoter is constitutive in *E. faecalis* - see ref. 1) to drive expression of the introduced ORF, and compare cell growth to control bacterial cells containing the parental vector lacking any introduced phage ORF. Recombinant plasmids are introduced into *E. faecalis* strain FA2-2 by electroporation, as previously described (Yamagishi, J., Kojima, T., Oyamada, Y., Fujimoto, K., Hattori, H., Nakamura, S., and Inoue, M. 1996. *Antimicrob. Agents Chemother.* 40, 1157-1163).

Cloning of ORFs with a Shine-Dalgarno sequence

ORFs with a Shine-Dalgarno sequence are selected for functional analysis of bacterial killing. Each ORF, from initiation codon to last codon (excluding the stop codon), will be amplified by PCR from phage genomic DNA. For PCR amplification of ORFs, each sense strand primer starts at the initiation codon and is preceded by a restriction site and each antisense strand starts at the last codon (excluding the stop codon) and is preceded by a different restriction site. The PCR product of each ORF will be gel purified and digested with the restriction enzymes with sites contained on

the PCR oligonucleotides. The digested PCR product is then gel purified using the Qiagen kit, ligated into the modified shuttle vector, and used to transform bacterial strain DH10 β . Recombinant clones are then picked and their insert sizes confirmed by PCR analysis using primers flanking the cloning site as well as restriction digestion.

The sequence fidelity of cloned ORFs will be verified by DNA sequencing using the same primers as used for PCR. In the cases that the verification of ORFs can not be achieved by one path of sequencing using primers flanking the cloning site internal primers will be selected and used for sequencing. Recombinant plasmids will be introduced into *E. faecalis* strain FA2-2 by electroporation, as previously described (Yamagishi, J., Kojima, T., Oyamada, Y., Fujimoto, K., Hattori, H., Nakamura, S., and Inoue, M. 1996. *Antimicrob. Agents Chemother.* 40, 1157-1163).

Induction of gene expression from the *ars* promoter.

If an inducible promoter is used, e.g., the *ars* promoter, induction can be assessed, for example, in either of the two methods.

1. Screening on agar plates

The functional identification of killer ORFs can be performed by spreading an aliquot of *E. faecalis* transformed cells containing phage 182 ORF onto agar plates containing different concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M). The plates are incubated overnight at 37°C, after which a growth inhibition of the ORF transformants on plates that contain arsenite are compared to plates without arsenite.

2. Quantification of growth inhibition in liquid medium

Cells containing different recombinant plasmids can be grown for overnight at 37°C in LB medium supplemented with the appropriate antibiotic selection. These are then diluted to the mid log phase ($OD_{540}=0.2$) with fresh media containing antibiotic and transferred to 96-well microtitration plates (100 μ l/well). Inducer is then added at different final concentrations (ranging from 2.5 to 10 μ M) and the culture incubated for an additional 2 h at 37°C. The effect of expression of the phage 182 ORFs on bacterial cell growth is then monitored by measuring the OD_{540} and comparing the rate of growth to the culture not containing inducer. As positive controls for growth inhibition, the *kiLA* gene of phage lambda (Reisinger, GR., Rietsch, A., Lubitz, W. and Blasi, U. 1993 *Virology* #193: 1033-1036), and the *holin/lysin* genes of the *Staphylococcus aureus* phage Twort (Loessner, MJ., Gaeng, S., Wendlinger, G.,

Maier, SK. and Scherer, S. 1998. *FEMS Microbiology Letters* #162:265-274) were subcloned into the *ars* inducible vector. An aliquot of the induced and uninduced culture can also be plated out on agar plates containing an appropriate antibiotic selection but lacking inducer. Following incubation overnight at 37°C, the number of colonies is counted. Any ORF showing bacteriostatic activity will show a lower, but detectable, number of colonies on the agar plates when grown in the presence of inducer as compared to when grown in the absence of inducer. Any ORF showing bacteriocidal activity will show no colonies on the agar plates, when grown in the presence of inducer as compared to when grown in the absence of inducer.

REFERENCES

1. Cohen, M.L. (1992). *Science* 257, 1050-1055.
2. Tenover, F.C. and McGowan Jr., J.E. (1998). Bacterial Infections of Humans, Epidemiology and Control. (A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 83-93.
3. Rusterholtz, K., and Pohlschroder, M. (1999). *Cell* 96, 469-470.
4. Neu, H.C. (1992). *Science* 257, 1064-1073.
5. Murray, B.E. (1990). *Clin. Microbiol. Rev.* 3, 46-65.
6. Gray, B.M. (1998). Bacterial Infections of Humans, Epidemiology and Control. (A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 673 - 711.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: A laboratory Manual*. Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press.
7. Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J.
8. Rost B,l and Sander C. (1996). *Ann. Rev. Biophy. Biomol. Struct.* 25, 113-136.
9. Garvey, K.J., Saedi, M.S., and Ito, J. (1985). *Gene* 40, 311-316.
10. Pickett, G.G. and Peabody, D.S. (1993). *Nucl. Acids Res.* 21, 4621-4626.
11. Gutiérrez, J., Vinos, J., Prieto, I., Mendez, E., Hermoso, J., and Salas, M. (1986). *Virology* 155, 474-483.
12. Yoshikawa, H., Garvey, K.J., and Ito, J. (1985). *Gene* 37, 125-130.
13. Martin, A.C., Lopez, R., Garcia, P. (1998). *J Bacteriol* 180, 210-217.

14. Steiner, M., Lubitz, W., Blasi, U. (1993). *J. Bacteriol.* **175**, 1038-1042.
- Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). *Genes Dev.* **7**, 555-569.
- Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). *Mol. Cell Biol.* **18**, 2697-2711.
- Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998). *Genes, Chromosomes & Cancer* **21**, 217-222.
- Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. (1997). *Nature* **387**, 921-924.
- Karimova, G., Pidoux, J., Ullmann, A., Ladant, D. (1998) *Proc. Natl. Acad. Sci.* **95**, 5752-5756.
- Sopta, M., Carthew, R.W., and Greenblatt, J. (1995) *J. Biol. Chem.* **260**, 10353-10369.
- Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997). *Anal. Chem.* **69**, 3995-4001.
- Swanström, M. and Adams, M.H. (1951). *Proc. Soc. Exptl. Biol. Med.* **78**, 372-375.

Example 15. Growth of *Streptococcus* bacteriophage Dp-1 and purification of genomic DNA.

The *Streptococcus pneumoniae* R6 propagating strain (PS) (Tomasz, 1966) was used as host to propagate its respective phage Dp-1 (McDonnell et al., 1975). (Alternatively, *Streptococcus (Diplococcus) pneumoniae* R36A could be used. Strain R36A is available from ATCC as #11733 or 27336. *Streptococcus pneumoniae* is also available from Felix d'Herelle Reference Center in Quebec, Canada as catalog number HER 1054. Other *S. pneumoniae* strains are also available from ATCC.) Two rounds of plaque purification of phage Dp-1 were performed on soft agar essentially as described in Sambrook et al. (1989). Briefly, the *Streptococcus* R6 PS strain was grown overnight at 37°C in K-Cat media [K-Cat: 10 g Bacto casitone, 5 g Bacto tryptone, 1 g Yeast extract, 5g Potassium chloride, 0.2% Glucose, 30mM Potassium phosphate buffer [pH 8] and 250,000 Units Catalase per liter (Boehringer Mannheim #10683600). The culture was then diluted 20 fold in K-CAT and

incubated at 37°C until the $OD_{540} = 0.2$ (early log phase) with constant agitation. In order to obtain single plaques, Dp-1 phage was subjected to 10-fold serial dilutions using the phage buffer (100 mM Tris-HCl [pH 7.5], 100 mM NaCl and 10 mM $MgCl_2$) and 10 μ l of each dilution was used to infect 0.5 ml of the cell suspension.

- 5 After incubation of 15 min at 37°C, 2 ml of melted soft agar (K-CAT supplemented with 0.8% of agar) were added to the mixture and poured onto the surface of 100 mm K-CAT agar plates [K-CAT supplemented with 1.2 % of agar]. After solidification of the soft agar layer, an additional 5 ml of melted soft agar was added to visualize distinct plaques (Ronda et al., 1978). After overnight incubation at 37°C, a single plaque was isolated, resuspended in 1 ml of phage buffer by end over end rotation for 2 hrs at room temperature, and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 37°C, a single plaque was isolated and used as a stock for all subsequent manipulations.

- The propagation procedure for bacteriophage Dp-1 was modified from the agar layer method of Swanstörn and Adams (1951). Briefly, the R6 strain of *Streptococcus pneumoniae* was grown to stationary phase overnight at 37°C in K-CAT. The culture was then diluted 20 fold in K-CAT and incubated at 37°C until the $OD_{540} = 0.2$. The suspension (15×10^7 Bacteria) was then mixed with 15×10^5 plaque forming units (pfu) to give a ratio of 100-bacteria/pfu. After incubation of 15 min at 37°C, 7.5 ml of melted soft agar (K-CAT plus 0.8% agar) were added to the mixture and poured onto the surface of 150 mm K-CAT agar plates and incubated 16 hrs at 37°C. After solidification of the soft agar layer, 7.5 ml of melted soft agar were added to each plate. To collect the plate lysate, 20 ml of K-CAT media were added to each plate and the soft agar layers were collected by scrapping off with a clean microscope slide followed by vigorous shaking of the agar suspension for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 rpm (2,830 xg) using a JA-10 rotor (Beckman) and the supernatant (lysate) was collected and subjected to a treatment with 10 μ g /ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, the phage suspension was adjusted to 10% (w/v) of PEG 8000 and 0.5 M of NaCl followed by incubation at 4°C for 16 hrs. The phage was recovered by centrifugation at 4,000 rpm (3,500 xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman). The pellet was resuspended with 2 ml of phage buffer (100 mM Tris-HCl [pH 7.5], 100 mM NaCl and 10 mM $MgCl_2$). The phage suspension was extracted with 1 volume of chloroform and further purified by centrifugation on a cesium chloride step gradient as described in Sambrook et al. (1989), using a TLS-55 rotor and centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 hrs at 28,000 rpm (67,000 xg) at 4°C. Banded phage was collected and ultracentrifuged again on an

isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000 xg) for 24 hrs at 4°C using a TLV rotor (Beckman). The phage was harvested and dialyzed for 4 hrs at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 µg/ml Proteinase K and 0.5% SDS and incubating for 1 hr at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris-HCl [pH 8.0], 1mM EDTA).

Example 16. DNA sequencing of the Bacteriophage Dp-1 genome.

Four micrograms of phage DNA was diluted in 200 µl of TE (10 mM Tris, [pH 8.0], 1 mM EDTA) in a 1.5 ml eppendorf tube and sonication was performed (550 Sonic Dismembrator, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 sec spaced by 15 sec cooling in ice/water for 3 to 4 cycles. The sonicated DNA was then size fractionated by electrophoresis on 1% agarose gels utilizing TAE (1 x TAE is: 40 mM Tris-acetate, 1 mM EDTA [pH 8.0]) as the running buffer. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen), with a final elution of 50 µl of 1 mM Tris [pH 8.5].

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and the Klenow fragment of *E. coli* DNA polymerase I, as follows. Reactions were performed in a reaction mixture (final volume, 100 µl) containing sonicated phage DNA, 10 mM Tris-HCl [pH 8.0], 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 50 µg/ml BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of the Klenow large fragment of DNA polymerase I (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the DNA was precipitated with ethanol and the final DNA pellet resuspended in 20 µl of H₂O.

Blunt-ended DNA fragments were cloned by ligation directly into the *Hinc* II site of the pKSII+ vector (New England Biolabs) dephosphorylated by treatment with calf intestinal alkaline phosphatase (New England Biolabs). A typical ligation reaction contained 100 ng of vector DNA, 2 to 5 µl of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 µl containing 800 units of T4 DNA ligase (New England Biolabs) and was incubated overnight at 16°C. Transformation and selection

of bacterial clones containing recombinant plasmids was performed in *E. coli* DH10 β according to standard procedures (Sambrook *et al.*, 1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 μ l LB and 100 μ g/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers flanking the *Hinc* II cloning site of the pKS vector. PCR amplification of the potential foreign inserts was performed in a 15 μ l reaction volume containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 μ M primer, 187.5 μ M each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 58°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using the QIAprep™ spin miniprep kit (Qiagen).

The nucleotide sequence of the extremities of each recombinant clone was determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism Big Dye™ primer cycle sequencing (21M13 primer: #403055)(M13REV primer: #403056) or ABI prism Big Dye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). To ensure co-linearity of the sequence data and the genome, all regions of the phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism Big Dye™ terminator cycle sequencing ready reaction kit.

Example 17. Bioinformatic management of primary nucleotide sequence.

Sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of the contigs. Phage DNA was used directly as sequencing template employing ABI prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). The complete sequence of *Streptococcus* bacteriophage Dp-1 is shown in Table 28.

A software program was used on the assembled sequence of bacteriophage Dp-1 to identify all putative ORFs larger than 33 codons. The software scans the primary nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG,

GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI(<http://www.ncbi.nlm.nih.gov/hibin-post/Taxonomy/wprintgc?mode=c>) for the bacterial genetic code. When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence. The predicted ORFs for bacteriophage Dp-1 are listed in Tables 29 and 30, and Fig. 6.

Sequence homology searches for each ORF were carried out using an implementation of BLAST programs. Downloaded public databases used for sequence analysis include:

- (i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) staphylococcus aureus NCTC 8325
- (<ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa>);
- vi) streptococcus pyogenes
- (ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- vii) PRODOM
- (ftp://ftp.toulouse.inra.fr/pub/prodom/current_release/prodom99.1.forblast.gz);
- viii) DOMO (<ftp://ftp.infobiogen.fr/pub/db/domo/>);
- ix) TREMBL (ftp://www.expasy.ch/databases/sp_tr_nrd/fasta/)

The results of the homology searches performed on the ORFs of bacteriophage Dp-1 are shown in Table 31.

Example 18. Sub-Cloning of Bacteriophage Dp-1 ORFs.

Preparation of the shuttle expression vector

Expression preferably utilizes a shuttle expression vector which is arranged such that expression of the exogenous bacteriophage Dp-1 ORF sequence is inducible. For example, the plasmid pLSE4 replicates in *E. coli*, and *S. pneumoniae* (Diaz and Garcia, 1990). This plasmid can be modified by conventional techniques to insert the inducible arsenite promoter, derived from the shuttle vector pT0021, in which the

firefly luciferase (*lucFF*) expression is controlled by the *ars* promoter/operator from a *S. aureus* plasmid (Tauriainen, S., Karp, M., Chang, W and Virta, M. (1997). Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite. *Appl. Environ. Microbiol.* 63:4456-4461). This modified shuttle vector will contain the *ars* promoter, *arsR* gene and a cloning site for introduction of individual phage ORFs downstream from a shine-dalgarno sequence.

Other inducible regulatory sequences can be utilized instead of the arsenite-inducible system. An example is a nisin-inducible system. The *nisA* promoter activity is dependent on the proteins NisR and NisK, which constitute a two-component signal transduction system that responds to the extracellular inducer nisin. The nisin sensitivity and inducer concentration required for maximal induction varies among the strains, but is functional in *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Enterococcus faecalis*, and *Bacillus subtilis*. Significant induction of the *nisA* promoter (10- to 60-fold induction) can be obtained in all of the species. A vector containing this promoter was published as Eichenbaum Z, Federle MJ, Marra D, de Vos WM, Kuipers OP, Kleerebezem M, and Scott JR (1998) *Appl Environ Microbiol* 64, 2763-2769. Other vectors, e.g., plasmids, can also be utilized which will allow replication and transcription in *Streptococcus*.

Alternatively, a constitutive promoter can be used to drive expression of the introduced ORF, and compare cell growth to control bacterial cells containing the parental vector lacking any introduced phage ORF. Recombinant plasmids are introduced into *S. pneumoniae* R6 as previously described (Diaz and Garcia, 1990)

Cloning of ORFs with a Shine-Dalgarno sequence

ORFs with a Shine-Dalgarno sequence are selected for functional analysis of bacterial killing. Each ORF, from initiation codon to last codon (excluding the stop codon), will be amplified by PCR from phage genomic DNA. For PCR amplification of ORFs, each sense strand primer starts at the initiation codon and is preceded by a restriction site and each antisense strand starts at the last codon (excluding the stop codon) and is preceded by a different restriction site. The PCR product of each ORF will be gel purified and digested with the restriction enzymes with sites contained on the PCR oligonucleotides. The digested PCR product is then gel purified using the Qiagen kit, ligated into the modified shuttle vector, and used to transform bacterial strain DH10 β . Recombinant clones are then picked and their insert sizes confirmed by PCR analysis using primers flanking the cloning site as well as restriction digestion. The sequence fidelity of cloned ORFs will be verified by DNA sequencing using the same primers as used for PCR. In the cases that the verification of ORFs can not be achieved by one path of sequencing using primers flanking the cloning site

internal primers will be selected and used for sequencing. Recombinant plasmids will be introduced into *S. pneumoniae* R6 as previously described (Diaz and Garcia, 1990).

Induction of gene expression from the *ars* promoter.

If an inducible promoter is used, e.g., the *ars* promoter, induction can be assessed, for example, in either of the two methods.

1. Screening on agar plates

The functional identification of killer ORFs can be performed by spreading an aliquot of *S. pneumoniae* transformed cells containing phage Dp-1 ORFs onto agar plates containing different concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M).

The plates are incubated overnight at 37°C, after which a growth inhibition of the ORF transformants on plates that contain arsenite are compared to plates without arsenite.

2. Quantification of growth inhibition in liquid medium

Cells containing different recombinant plasmids can be grown for overnight at 37°C in LB medium supplemented with the appropriate antibiotic selection. These are then diluted to the mid log phase ($OD_{540}=2$) with fresh media containing antibiotic and transferred to 96-well microtitration plates (100 μ l/well). Inducer is then added at different final concentrations (ranging from 2.5 to 10 μ M) and the culture incubated for an additional 2 hrs at 37°C. The effect of expression of the phage Dp-1 ORFs on bacterial cell growth is then monitored by measuring the OD_{540} and comparing the rate of growth to the culture not containing inducer. [As positive controls for growth inhibition, the *kilA* gene of phage lambda (Reisinger, GR., Rietsch, A., Lubitz, W. and Blasi, U. 1993 *Virology* #193: 1033-1036), and the *holin/lysin* genes of the *Staphylococcus aureus* phage Twort (Loessner, MJ., Gaeng, S., Wendlinger, G., Maier, SK. and Scherer, S. 1998. *FEMS Microbiology Letters* #162:265-274) can be subcloned into the *ars* inducible vector. An aliquot of the induced and uninduced culture can also be plated out on agar plates containing an appropriate antibiotic selection but lacking inducer. Following incubation overnight at 37°C, the number of colonies is counted. Any ORF showing bacteriostatic activity will show a lower, but detectable, number of colonies on the agar plates when grown in the presence of inducer as compared to when grown in the absence of inducer. Any ORF showing full bacteriocidal activity will show no colonies on the agar plates, when grown in the presence of inducer as compared to when grown in the absence of inducer.

REFERENCES

15. Cohen, M.L. (1992) *Science* 257, 1050-1055.

16. Tenover, F.C. and McGowan Jr., J.E. (1998) Bacterial Infections of Humans. Epidemiology and Control, (A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 83-93.
17. Rusterholtz, K., and Pohlschroder, M. (1999) *Cell* **96**, 469-470.
18. Klugman, K.P. (1990) *Clin. Microbiol. Rev.* **3**, 171-196.
19. Fenoll, A., Martin Bourgon, C., Munoz, R., Vicioso, D., Casal, J. (1991) *Rev. Infect. Disease* **13**, 56-60.
20. Jorgensen, J.H., Doern, G. V., Maher, L. A., Howell, A. W., Redding, J. S. (1990) *Antimicrob. Agents Chemother.* **34**, 2075-2080.
21. Neu, H.C. (1992) *Science* **257**, 1064-1073.
- Hsueh, P. R., Wu, J. J., Hsiue, T. R. (1996) *J Formos Med Assoc* **5**, 364-371.
 - Garcia, P., Martin, A.C., and Lopez, R. (1997) *Microbial Drug Res.* **3**, 165-176.
 - Martin, A.C., Lopez, R., and Garcia, P. (1996) *J. Virol.* **70**, 3678-3687.
 - Sheehan, M.M., Garcia, J.L., Lopez, R., and Garcia, P. (1997) *Mol. Microbiol.* **25**, 717-725.
 - Kodaira, M., Biswas, S.B., and Kornberg, A. (1983) *Mol. Gen. Genet.* **192**, 80-96.
 - Maki, S. and Kornberg, A. (1988) *J. Biol. Chem.* **263**, 6547-6554.
 - Tsuchihashi Z, Kornberg A. (1990) *Proc. Natl. Acad. Sci. USA.* **87**, 2516-2520.
 - Lee, S.H. and Walker, J.R. (1987) *Proc Natl Acad Sci USA* **84**, 2713-2717.
 - Smidt, C.R., Steinberg, F.M., Rucker, R. (1991) *Proc Soc Exp Biol Med* **197**, 19-26.
 - Frank, D.W., (1997) *Mol Microbiol.* **26**, 621-629.
 - Nardese, V., Gutlich, M., Brambilla, A., Carbone, M.L. (1996) *Biochem Biophys Res Commun* **218**, 273-279.
 - Mancini, R., Saracino, F., Buscemi, G., Fischer, M., Schramek, N., Bracher, A., Bacher, A., Gutlich, M., Carbone, M.L. (1999) *Biochem Biophys Res Commun* **255**, 521-527.
 - Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: A laboratory Manual*. Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press.

22. Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J.
23. Rost B.I and Sander C. (1996) *Ann. Rev. Biophys. Biomol. Struct.* **25**, 113-136.
24. Garvey, K.J., Saedi, M.S., and Ito, J. (1985) *Gene* **40**, 311-316.
25. Pickett, G.G. and Peabody, D.S. (1993) *Nucl. Acids Res.* **21**, 4621-4626.
26. Gutiérrez, J., Vinos, J., Prieto, I., Mendez, E., Hermoso, J., and Salas, M. (1986) *Virology* **155**, 474-483.
27. Yoshikawa, H., Garvey, K.J., and Ito, J. (1985) *Gene* **37**, 125-130.
28. Martin, A.C., Lopez, R., Garcia, P. (1998) *J Bacteriol* **180**, 210-217.
29. Steiner, M., Lubitz, W., Blasi, U. (1993) *J. Bacteriol.* **175**, 1038-1042.
- Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). *Genes Dev.* **7**, 555-569.
 - Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998) *Mol Cell Biol.* **18**, 2697-2711.
 - Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998) *Genes, Chromosomes & Cancer* **21**, 217-222.
 - Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. (1997) *Nature* **387**, 921-924.
 - Karimova, G., Pidoux, J., Ullmann, A., Ladant, D. (1998) *Proc. Natl. Acad. Sci.* **95**, 5752-5756.
 - Söpta, M., Carthew, R.W., and Greenblatt, J. (1995) *J. Biol. Chem.* **260**, 10353-10369.
 - Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997) *Anal. Chem.* **69**, 3995-4001.
 - Tomasz, A. (1966) *Journal of Bacteriology* **91**, 1050-1061.
 - McDonnell, M., Ronda, LC and Tomasz, A. (1975) *Virology* **63**, 577-582.
 - Ronda C., Lopez, R., Tomasz, A. and Portoles A. (1978) **26**, 221-225.

- Swanström, M. and Adams, M.H. (1951) Proc. Soc. Exptl. Biol. Med. 78, 372-375.
- Diaz E and Garcia JL. (1990) Gene 90, 163-167.

- Tauriainen, S., Karp, M., Chang, W and Virta, M. (1997). Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite. Appl. Environ. Microbiol. 63:4456-4461.

All patents and publications mentioned in the specification are indicative of the levels of skill of those skilled in the art to which the invention pertains. All references cited in this disclosure are incorporated by reference to the same extent as if each reference had been incorporated by reference in its entirety individually.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The specific methods and compositions described herein as presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. For example, those skilled in the art will recognize that the invention may suitably be practiced using a variety of different bacteria, bacteriophage, and sequencing methods within the general descriptions provided.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is not intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

5 In addition, where features or aspects of the invention are described in terms of
Markush groups or other grouping of alternatives, those skilled in the art will
recognize that the invention is also thereby described in terms of any individual
10 member or subgroup of members of the Markush group or other group. For example,
5 if there are alternatives A, B, and C, all of the following possibilities are included: A
separately, B separately, C separately, A and B, A and C, B and C, and A and B and
C. Thus, for example, for the bacteria and phage specified herein, the embodiments
15 expressly include any subset or subgroup of those bacteria and/or phage. While each
such subset or subgroup could be listed separately, for the sake of brevity, such a
10 listing is replaced by the present description.

Thus, additional embodiments are within the scope of the invention and within
20 the following claims.

Table 1

Phages against human and animal pathogenic bacteria

I. Pathogen name	Phage name	II. Catalog#	Origin/reference
<i>Acinetobacter calcoaceticus</i>	A3/2 A10/45 A36 B9GP B ₉ PP BS46 E13 E14 S31		Felix d'Herelle Reference Centre, Quebec, Quebec
	Ap3 P78		J. Bacteriol 1984. 157: 179-183 J. Gen. Microbiol 1986.132: 2633-2636
<i>Acinetobacter haemolyticus</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Acinetobacter johnsonii</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Acinetobacter sp.</i>	BP1		J. Virol. 1968.2:716-722
	G4, HP2, HP3 & HP4		Can. J. Microbiol. 1966.12:1023-1030 & J. Virol. 1974.13:46-52 & Arch. Virol. 1994.135:345-354
	A1, A4, A9 & 196		Arch. Virol. 1994.135:345-354
	HP1		Can. J. Microbiol. 1966.12:1023-1030
	A19, A23, A29, A31, A33, A34, A3759 & 2845		J. Microsc (Paris) 1973.16:215-224 & CR. Hebdo Seances Acad. Sci. Ser D. Sci Natur (Paris) 278:1907-1909 & Arch. Virol. 1994.135:345-354 & Rev. Can. Biol. 1970.29:317-320
<i>Actinobacillus actinomycescomitans</i>			FEMS Microbiol Lett 1994. 119:329-337

			Infect. Immun. 1982. 35: 343-349
			Mol. Gen. Genet. 1998. 258: 323-325
	Aap247		Oral Microbiol. Immunol. 1997. 12: 40-46
<i>Actinomyces viscosus</i>		43146-B1	The American Type Culture Collection
			Infect. Immun. 1985. 48: 228-233
			Infect. Immun. 1988. 56: 54-59
			Plasmid 1997. 37: 141-153
<i>Aeromonas hydrophila</i>	PM2** & PM3		FEMS Microbiol. Lett. 1990. 57: 277-282
	Aeh1 Aeh2 PM4 PM5 PM6 T7-ah		Felix d'Herelle Reference Centre, Quebec, Quebec

55

<i>Bordetella paraperussis</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
			Mol. Gen. Mikrobiol. Virusol. 1988.4: 22-25
			Zh.Mikrobiol.Epidemiol.Immuno. 1987.5:9-13
	41405		Zh.Mikrobiol.Epidemiol.Immuno. 1987.5:9-13
<i>Brucella abortus</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
	10/I 24/II 212/XV	23448-B1 23448-B2 23448-B3 17385-B1 17385-B2	The American Type Culture Collection
	BK-2, TB & Fj**		Zh.Mikrobiol.Epidemiol.Immunobiol.1983.2: 48-52
	R/c & R/O		Dev. Biol. Stand. 1984.56: 55-62
<i>Brucella canis</i>	R/c		Dev. Biol. Stand. 1984.56: 55-62
<i>Brucella melitensis</i>	BK-2	23456-B1	The American Type Culture Collection
<i>Brucella suis</i>	Wb		Zentralbl.Veterinarmed.1975.22:866-867

	Fi** & TB		Zh.Mikrobiol.Epidemiol.Immunobiol.1983.2: 48-52
<i>Brucella sp.</i>			Can. J. Vet. Res. 1989.53: 319-325
			Res. Vet. Sci. 1988. 44: 45-49
	R		Zh.Mikrobiol.Epidemiol.Immunobiol.1983.2: 48
<i>Campylobacter coli</i>		43133-B1 43134-B1	The American Type Culture Collection
<i>Campylobacter coli</i> (Cont'd)	18 19 20	43135-B1 43136-B1	The American Type Culture Collection
<i>Campylobacter jejuni</i>	1 2 3 4 5 6 7 8 9 10 11 12 13 14 17 18 19 20	35918-B1 35919-B1 35920-B1 35921-B1 35918-B2 35920-B2 35922-B2 35923-B1 35924-B1 35925-B1 35925-B2 35922-B2 35924-B2 35922-B3 43133-B1 43134-B1 43135-B1 43136-B1	The American Type Culture Collection
<i>Campylobacter</i> (<i>Helicobacter</i>) <i>pylori</i>	HP1		J. Med. Microbiol.1993. 38: 245-249
<i>Chlamydia psittaci</i>	Chp1**		J. Gen. Virol. 1989. 70: 3381-3390
<i>Clostridium</i> <i>acetobutylicum</i>	CAK-1		J.Bacteriol.1993.175:3838-3843

<i>Clostridium botulinum</i>			Nucleic Acids Res.1990.18:1291
			Bioch.Biophys.res.Commun.1990.171.1304-1311
			Microbiol.immunol.1981.25:915-927
			J.Vet.Med.Sci.1992.54:675-684
	CE β & CE γ		
<i>Clostridium difficile</i>	41 & 56		J. Clin.Microbiol. 1985.21:251-254

<i>Clostridium perfringens</i>			Rev.Can.Biol.1977.36:205-215
			FEMS Microbiol.Lett. 1990.54:323-326
<i>Clostridium sporogenes</i>	59 70 71 72S 72L	8074-B1 17886-B1 17886-B3 17886-B4 17886-B5 17886-B6	The American Type Culture Collection
<i>Clostridium tetani</i>	A & B		Rev.Can.Biol.1978.37:43-46
<i>Corynebacterium diphtheriae</i>			Vopr.Virusol.1986.31:577-584
<i>Corynebacterium pseudotuberculosis</i>	NN	12319-B1	The American Type Culture Collection
<i>Corynebacterium sp</i>	DLC 2921/49	12052-B1	The American Type Culture Collection

5

10

15

20

25

30

35

40

45

50

55

<i>Enterococcus faecalis</i>	42	19948-B1	The American Type Culture Collection
<i>Enterococcus faecium</i>	124 133	19950-B1 19953-b2 19953-B1	The American Type Culture Collection

5

10

15

20

25

30

35

40

45

50

55

<i>Escherichia coli</i>		11303-B14	The American Type Culture Collection
		11303-B10	
		11303-B21	
		8677-B1	
		11303-B13	
		13706-B4	
<i>Escherichia coli</i> (Cont'd)		15766-B1	The American Type Culture Collection
		15766-B1	
		1242-B5	
		15669-B2	
		15767-B1	
		11303-B16	
		27-65-B1	
		25065-B2	
	C204	15669-B1	
	E1	15597-B1	
	f1**	21816-B1	
	f2**	23724-B9	
	FCZ	15593-B1	
	fd**	25404-B1	
		29746-B1	
		23631-B1	
		25868-B1	
		25298-B1	
		25298-B2	
		11303-B37	
		11303-B24	
		11303-B26	
		11303-B27	
		11303-B28	
		11303-B29	
		11303-B30	
		11303-B33	
		11303-B31	
		11303-B25	
		11303-B35	
		11303-B34	
		11303-B36	
		11303-B32	
		13706-B5	
		11303-B1	
		11303-B2	
		11303-B3	
		11303-B4	
		35060-B1	
		35060-B2	
		35060-B3	
		11303-B5	
		11303-B6	
		11303-B7	
		11303-B38	
		12141-B1	

<i>Escherichia coli</i> (Cont'd)			The American Type Culture Collection
		11303-B20	
		11303-B17	
		11303-B15	
		11303-B11	
547		11303-B18	
UV1		13706-B2	
UV47		23724-B2	
UV375		23724-B1	
$\alpha 3^{**}$		23724-B3	
λ **		23724-B4	
λ C-17		23724-B5	
λ sus P-3		23724-B6	
λ sus R-5		23724-B7	
λ sus J-6		23724-B8	
λ sus O-8		35860-B1	
λ sus A-11		13706-B3	
λ ind'		15597-B2	
$\phi 92$		13706-B1	
ϕR		49696-B1	
$\phi V-1$			
$\phi X174^{**}$			
$\phi Xcs70am-3$			
G4** & K**			Biochim.Biophysica Acta.1992.1130:277-288
BF23**			J.Bacteriol.1977.129:265-275
Mu1			J.Ultrastruct.Res.1966.14:441-448
Hp17			J.Mol.Biol.1991.218:705-721
K3** & Ox2**			FEBS Lett.1987.215:145-150
Rb18**, Rb51 & Rb69**			J.Bacteriol.1990.172:180-186
H1**, H3, H8, K9, K18 & Ox1			Mol.Gen.Genet.1990.221:491-494
M1**, Tula** & Tulb**			J.Mol.Biol.1987.196:165-174
K10			J.Bacteriol.1979.140:680-686
Qsr'			J.Bacteriol.1985.162:256-262
B278			J.Gen.Microbiol.1988.134:1333-1338
ϕ i 80**			FEMS Microbiol.Lett.1994.119:71-76
ϕ i m173			Genetika 1985.21:673-675
tf-1			J.Gen.Microbiol.1987.133:953-960
P4 & ϕ iR73			Mol.Microbiol.1995.18:201-208
I ₂ -2			J.Gen.Microbiol.1982.128:2797-2804
PRD1			Virology 1990.177:445-451
K3hx			Mol.Gen.Genet.1987.206:110-115
933J** & 933W**			Infect.Immunity.1986.53:135-140
H19-B**			J.Bacteriol.1987.169:4308-4312
Tcp-111			Zentralbln. Bakteriolo. Mikrobiol. Hyg. 1988.270: 41-51

	N4**	Vet.Microbiol.1992.30:203-212
	Phi 80 up	Ann.Inst.Pasteur.1971.120:121-125
	Obeta I	J.Bacteriol.1978.133:172-177
10	PICM	J.Gen.Microbiol.1978.107:73-83
	PA-2**	J.Bacteriol.1990.172:1660-1662
	186**	Mol.Gen.Genet.1982.187:87-95
	186.IX.B	Mol.Microbiol.1992.6:2629-2642
	21**	Virology 1983.129:484-489
	P4**	Microbiol.Rev.1993.57:683-702
15	82**	J.Biol.Chem.1987.262:11721-11725
	PSP3	J.Bacteriol.1996.178:5668-5675
	HK022**	Nucleic Acids Res.1994.22:354-356
	D108**	Nucleic Acids Res.1986.14:3813-3825
	<i>Escherichia coli</i> (Cont'd)	
20	Rb49	J.Mol.Biol.1997.267:237-249
	Ikc**	J.Mol.Biol.1985.181:27-39
	P22dis	Mol.Gen.Genet.1978.166:233-243
	N15**	J.Bacteriol.1996.178:1484-1486
	IF**	Proc.R.Soc.Lond.B.Biol.Sci.1991.245:23-30
	Stx2Phi-I & Stx2Phi-II	Infect.Immun.1998.66:4100-4107
25	18	Virology 1987.156:122-126
	X	J.Gen.Microbiol.1981.126:389-396
	AC3	Mol.Microbiol.1991.5:715-725

	BW-1 C-1 E920g Esc-7-11 H19J Haiti HK243 Ia K20 K30 KL, M Mu** O103 O157:H7 P1D pt1 PilH α PR64FS PR772 SS4 β 4Q λ vir** Ω 8 O9-1 92		Felix d'Herelle Reference Centre, Quebec, Quebec
	<i>Haemophilus influenzae</i>	HP1**	Nucleic Acids Res. 1996.24:2360-2368
		S2**	Gene 1997. 196: 139-144
	<i>Halobacterium cutirubrum</i>	S45	Felix d'Herelle Reference Centre, Quebec, Quebec
	<i>Halobacterium halobium</i>		Felix d'Herelle Reference Centre, Quebec, Quebec
			Can.J.Microbiol.1982.28:916-921
	<i>Halobacterium salinarum</i>		Biol.Chem.Hoppe Seyler 1994.375:747-757

<i>Klebsiella oxytoca</i>	tf-1		J.Gen.Microbiol.1987.133:953-960
<i>Klebsiella pneumoniae</i>	60	23356-B1	The American Type Culture Collection
	92	23357-B1	
	K19Q		Felix d'Herelle Reference Centre, Quebec, Quebec
	FC3-1 & FC3-9		Can.J.Microbiol.1991.37:270-275
	FC3-10		FEMS Microbiol.Lett.1991.67:291-297
<i>Klebsiella sp.</i>	K11**		Mol.Gen.Genet. 1990.221:283-286
<i>Leptospira sp.</i>	LE1, LE3 & LE4		Res.Microbiol.1990.141:1131-1138
<i>Listeria monocytogenes</i>	243	23074-B1	The American Type Culture Collection
	197,1313 & 9425		Appl.Environ.Microbiol.1997.63:3374-3377
	H387 & H387-A		Appl.Environ.Microbiol.1993.59:2914-2917
	5775,6223 &12682		APMIS.1993.101:160-167
	2389, 2671, 4211 & 2685		Intervirology 1994.37:31-35 & Zentralbl.Bakteriol.Mikrobiol.Hyg.1986.261:1 2-28
	4b, 4ab, 4g & 3c		Ann.Microbiol. (Paris) 1977.128:185-198
	A118, A500 & A511**		Mol.Microbiol. 1995.16:1231-1241-992
	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 19 & 20		Ann.Microbiol. (Paris) 1979.130B:179-189
	1/2a, 1/2b, 3c, 4ab, 6a & 6b		Clin.Invest.Med.1984.7:229-232
	φLMUP35 2685		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Listeria innocua</i>	4211		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Micrococcus luteus</i>		4698-B1	The American Type Culture Collection
		4698-B4	
	N3	4698-2	
	N4	4698-B3	
	N8		
<i>Micrococcus luteus</i>	N17		Can.J.Microbiol. 1979.25:1027-1035
<i>Mycobacterium smegmatis</i>	BK-3	27203-B1	The American Type Culture Collection
	Bo1**	27204-B1	
	Bo 6	27205-B1	
	Bo 6II	27205-B2	
	Bo 6III	27205-B3	
	Mc-2	607-B6	
	Mc-4	607-B7	
	NN	11727-B1	
	Phagus lacticola	11759-B1	
	R1	607-B1	

Legendre Leo Roy Sedge	HER 317 HER 330 HER 333 HER 335 HER 334 HER 331 HER 316	Felix d'Herelle Reference Centre, Quebec, Quebec
		Mol.Microbiol.1993.7:395-405
		J.Mol.Biol.1998.279:143-164
		Proc.Natl.Acad.Sci.USA.1988.84:2833-2837
		Mol.Biol.Rep. 1981.30:11-15
		Proc.Natl.Acad.Sci.USA 1997.94:10961-10966
	29M, 31M, 122, 154, 37, 29D, 46, 139, 110, 141, 74D, AG1 & DS6A	Arch.Virol.1993.133:39-49 & Am.Rev.Respir.Dis.1975.112:17-22
<i>Mycobacterium fortuitum</i>	Bo 4 Bo 7	23052-B1 27207-B1 27207-B2 The American Type Culture Collection

<i>Mycobacterium leprae</i>			Ann.Microbiol. (Paris) 1982.133:93-97
<i>Mycobacterium tuberculosis</i>	DS6A	25618-B1 25618-B2 4243-B1	The American Type Culture Collection
	110, 139 & 33D		Arch.Virol.1993.133:39-49
	AG1,GS4E, BG1, PH & BK1		The Biology of Mycobacteria.Academic Press,Toronto 1982 (Ratledge & Stanford) 1982.309-351
<i>Mycobacterium sp</i>	Phagus pellegrini NN B1	11760-B1 11761-B1 23239-B1	The American Type Collection Culture

	TM4, ph60, ph72, PhAE39, phAE40 & Bxb1		Microbiology 1995.141:1173-1181
	C2		Experientia 1969.25:1112-1113
	18 & 115		J.Gen.Virol.1987.68:949-956
	63		Gruzlica 1968.36:617-622
	phlei & butyricum		J.Gen.Virol.1975.29:235-238
	MyF3P-59a		Z.Allg.Mikrobiol.1968.8:29-37
	Bo2a		J.Gen.Virol.1973.20:75-87
	D4,D28 & D32		J.Exptl.Med.1966.123:327-340
	HC		J.Bacteriol.1963.86:608-609
	<i>Mycobacterium vaccae</i>	B5	15483-B1
	<i>Mycobacterium phlei</i>	NN Bo 2 Bo 2h Bo 3	11728-B1 11758-B1 27086-B2 27086-B1
	<i>Mycoplasma arthritidis</i>	MAV1**	Infect.Immunity.1995.63:4016-4023
	<i>Mycoplasma hyorhinis</i>	Hr-1	Arch.Virol.1983.77:81-85
	<i>Mycoplasma pneumoniae</i>	Br-1	Arch.Virol.1983.75:1-15
	<i>Mycoplasma pulmonis</i>		Plasmid 1995. 33: 41-49
	<i>Mycoplasma sp.</i>		J.Gen.Microbiol.1985.131:3117-3126
			J. Virol.1986.59:584-590
			Gene 1994. 141: 1-8

5

10

15

20

25

30

35

40

45

50

55

		Microbios 1990. 64: 111-125
		Infection & Immunity 1995. 63: 4016-4023
		Med. Biol. 1982. 60: 116-120
MV-L2 &		Arch. Virol. 1979. 61: 289-296
		Acta. Virol. 1978. 22: 443-450
		J. Gen. Virol. 1979. 42: 315-322
		Virology 1973. 55: 118-126

			Science 1971.173:725-727
<i>Neisseria perflava</i>			J.Clin.Microbiol.1976. 4:87-91
<i>Nocardia erythropolis</i>	φC		J.Gen.Virol.1974.23:247-254
	φEC		J.Bacteriol.1976.126:1104-1107
	B225		Arch.Exp.Veterinarmed.1981.35:433-436
	B939a		Am.J.Vet.Res.1978.39:1565-1566
<i>Pasteurella multocida</i>	Nos.115, 32, 967 & 1075		Vet.Med.Nauki. 1977.14:33-36
<i>Propionibacterium acnes</i>	NN	29399-B1	The American Type Collection Culture

<i>Pseudomonas aeruginosa</i>	2	12175-B1	The American Type Culture Collection
	2A	12175-B2	
	2B	12175-B3	
	11	12175-B4	
	16	14205-B1	
	24	14206-B1	
	27	14207-B1	
	44	14208-B1	
	73	14209-B1	
	95	14210-B1	
	109	14211-B1	
	113	14212-B1	
	249	14213-B1	
	B3	14214-B1	
	Hoff 2	15692-B1	
	Hoff 3	14203-B1	
	Pa	14204-B1	
	Pb	12055-B1	
	PB-1	12055-B2	
	Pc	15692-B3	
	Pf	12055-B3	
	PP7**	25102-B1	
		15692-B2	
			Felix d'Herelle Reference Centre, Quebec, Quebec
	7 & 31		
	PI3**		J.Virol.1983.47:221-223
	φ-MC		Can.J.Microbiol.1969.15:1179-1186
	PI1**		J.Mol.Biol.1991.218:349-364
	PR4**		J.Gen.Virol.1979.43:583-592
	A7		J.Bacteriol.1992.174:2407-2411
	KF1		J.Biochem.1983.93:61-71
	CTX**		Mol.Microbiol.1993.4:1703-1709
	12**		J.Virol.1977.24:135-141

5

10

15

20

25

30

35

40

45

50

55

	<p>φKZ, 21, φNZ, PMN17, PTB80, 68, PB-1, E79, 16, 109, 352, 1214, F8, 71, 337, M4, φC17, SL2, B17, Li-24, φmnP78, PS17**, φ1, 73, M6, Li-2, 7, φmnF82, PTB2, PTB20, PTB42, φKF77, 31, PTB21, 119x, φPLS27, B3, 258, Hw12, PM57, PM62, PM105, 148, PM681, 198, 218, 222, 242, 246, PC131, φC11, SL5, D3112**, Jb19, F7, PM69, PM13, PM61, PM113, φ240, 249 & 269</p>		dd
--	---	--	----

5

10

15

20

25

30

35

40

45

50

55

<i>Pseudomonas aeruginosa</i> (Cont'd)	297, 309, 318, 11,		Arch. Virol. 1993.131:141-151
---	-----------------------	--	-------------------------------

5

10

15

20

25

30

35

40

45

50

55

<i>Pseudomonas cepacia</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Pseudomonas fragi</i>		27362-B1 27363 B1	The American Type Culture Collection
<i>Pseudomonas phaseolicola</i>	wy ø6		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Pseudomonas putida</i>	gb-1	12633-B1	The American Type Culture Collection
<i>Pseudomonas syringae</i>	ø6	40492-B1 21781-B1	The American Type Culture Collection
<i>Pseudomonas sp.</i>	PPs-G3	49780-B1	The American Type Culture Collection
<i>Salmonella bareilly</i>	Sab 2		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Salmonella enteritidis</i>	1, 2, 3 & 6 2a, 3a, 4a, 5a, 6a, 7a, 8a, 9a, 15, 19, 20 & 21**		Epidemiol Infect. 1995.114:227-236 Vet. Med. Nauki. 1975.12:55-60
<i>Salmonella newington</i>	Epsilon 34		J. Struct. Biol. 1995.115:283-289
<i>Salmonella newport</i>		27869-B1 27869-B2	The American Type Culture Collection
	16-19		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Salmonella paratyphi</i>	Paratyphoid A Jersey	19940-B1 12176-B1	The American Type Culture Collection
<i>Salmonella senftenberg</i>	SasL1, SaL2, SaL3, SaL4, SaL5 & SaL6		Indian J. Med. Res. 1997.105:47-52
<i>Salmonella typhimurium</i>	P22** SL-1 MB78** SE1 LT2 ES18** L**	19585-B1 40282	The American Type Culture Collection J. Virol. 1982.41: 1038-1043 J. Gen. Microbiol. 1986.132:1035-1041 Virology 1971.45:835-636 Virology 1970.42:621-632 J. Virol. 1985.56:1034-1036

	P1CM ctr-100		Mol.Gen.Genet.1975.138:113-126
	F22		Genet.Res.1986.48:139-143
	Fels 1		J.Gen.Virol.1978.38:263-272
	Fels 2		Genet.Res.1986.48:139-143
	Px		Mol.Gen.Genet.1970.108:184-202
	P1kc		Virology 1974.60:503-514
	A3 & A4		J.Bacteriol. 1987.169:1003-1009
	HT		Genet.Res.1976.27:315-322
<i>Salmonella typhimurium</i> (Cont'd)	IRA		J.Basic Microbiol. 1990.30:707-716
	MudI		Mol.Gen.Genet. 1986.202:327-330
	P22 (cir4-1, cir5-1 & cir6-1)		Mol.Gen.Genet.1984.198:105-109
	BF23**		Mol.Gen.Genet.1976.147:195-202
	Kb1		J.Bacteriol.1974.117:907-908
	P221dis		J.Gen.Virol.1978.41:367-376
	PRD1**		Virology 1990.177:445-451
	I-2**		J.Gen.Microbiol.1982.128:2797-2804
	tf-1		J.Gen.Microbiol.1987.133:953-960
<i>Salmonella typhosa/typhi</i>	X**		J.Gen.Microbiol.1981.126:389-396
	8	19937-B1	The American Type Culture Collection
	23	19938-B1	
	25	19939-B1	
	46	19942-B1	
	53	19943-B1	
	163	19946-B1	
	175	19947-B1	
	VII	27870-B1	
	VIJ	27870-B2	
	O1		Felix d'Herelle Refrence Centre, Quebec, Quebec
	VIII		Chung Hua Liu Hsing Ping H.T.C.1992.13:288
	j2		J.Gen.Microbiol.1983.129:3395-33400
<i>Salmonella sp.</i>	P3	25957-B1	The American Type Culture Collection
	P4**	25957-B2	
	P9a	25957-B3	
	P9c	25957-B4	
	P10	25957-B5	
	102	19945-B1	
	Chi (χ)	9842-B1	
	R34	97541	
	MG40		Virology 1968.34:521-530
	P14		Microb.Pathog.1990.8:393-402
	PSP3		Virology 1992.188:414
	Ike**		Zentralbl.Bakteriol.1976.234:294-304
	P27 & 9NA		J.Virol.1986.12:921-931
<i>Sphaerophilus natans</i>	SN1		Appl.Environ.Microbiol.1979.37:1025-1030

<i>Shigella dysenteriae</i>	P2 p80	23351-B1 11456b 11456a-B1	The American Type Culture Collection
<i>Shigella flexneri</i>	D20	12661-B1	The American Type Culture Collection
	SfII**		Mol.Microbiol.1997.26:939-950
	SfV**		Gene 1997.22:217-227
	Sf6**		Mol.Microbiol.1995.18:201-208
	SfX		Gene 1993.129:99-101
<i>Shigella sonnei</i>	C16**		
	Ufa		Mol.Biol.(Mosk) 1977.11:323-331
<i>Shigella sp</i>	37	23354-B1	The American Type Culture Collection
<i>Spiroplasma citri</i>	SpV1		Plasmid 1993.29:193-205
<i>Spiroplasma sp.</i>	SpV1-R8A2B		Nucleic Acids Res. 1990.18:1293
	SpV3		Isr.J.Med.Sci.1987.23:429-433
	Sp V4		J.Bacteriol.1987.169:4950-4961
<i>Staphylococcus albus</i>			Staphylococci & Staphylococcal Infections.1997. Vol1:503-508 (Karger,Basel)

5

10

15

20

25

30

35

40

45

50

55

<i>Staphylococcus aureus</i>			The American Type Culture Collection
		27702-B1	
		27703-B1	
		27704-B1	
		23360-B1	
		23361-B1	
	15	27705-B1	
	17	27712-B1	
	29	27690-B1	
	42D**	27691-B1	
	42E	27692-B1	
	47	27693-B1	
	52	27694-B1	
	52A	27695-B1	
	53	27696-B1	
	54	27697-B1	
	55	27698-B1	
	71	27699-B1	
	75	27693-B2	
	77	27700-B1	
	79	27701-B1	
	80	27706-B1	
	81	27707-B1	
	83A	27708-B1	
	84	33742	
	85**	33741-B1	
	88	15565	
	92	19685-B1	
	5504'	11987-B1	
	K	11988-B1	
	PI	15752-B1	
	PI4		
	UC18		

		HER 101 HER 239 HER 283 HER 49	Felix d'Herelle Reference Centre, Quebec, Quebec
	Twon**		
	$\phi 11^{**}$		J.Bacteriol.1988.170:2409-2411
	$\phi 13^{**}$ & $\phi 42^{**}$		J.Gen.Microbiol.1989.135:1679-1697
	L54a**		J.Bacteriol.1986.166:385-391
	80a**		Can.J.Microbiol.1996.43:612-616
	94,95 & 96		J.Clin.Microbiol.1988.26:2395-2401
	$\phi 131, A_3$ & A_5		Staphylococci & Staphylococcal Infections.1997. Vol1:503-508 (Karger,Basel)
	Phi PVL**		Gene 1998.215:57-67
	<i>Staphylococcus carnosus</i>	BaSTC2	Felix d'Herelle Reference Centre, Quebec, Quebec
	<i>Staphylococcus epidermidis</i>	1a, 2b, 3a, 4b, 5a, 6b, 7b, 8c, 9a, 10a, 11b, 12a & 13b	Can.J.Microbiol.1988.34:1358-1361
		41, 63, 118II, 138, 245, 336, 392 & 550	Res.Virol.1994.145:111-121
	<i>Staphylococcus saprophyticus</i>	1154A, 1405, 1314, 1139 & 1259	Res.Virol.1990.141: 625-635 & Res.Virol.1994.145:111-121
	<i>Staphylococcus sp.</i>	Phi 812, Phi 131, SK311 & U16	Virology 1998.246:241-252
	<i>Streptococcus faecalis</i>	VD13	HER44 Felix d'Herelle Reference Centre, Quebec, Quebec
	<i>Streptococcus faecium</i>	PE1	Zentralbl.Bakteriol.1975.231:421-425
	<i>Streptococcus oralis</i>	Cp-1** & Cp- 7**	FEMS Microbiol.Lett.1989.65:187-192

<i>Streptococcus pneumoniae</i>	Cp-1**	HER223	Felix d'Herelle Reference Centre, Quebec, Quebec
	Cp-1**, Cp-5**, Cp-7**, Cp-9**, w-1 & w-2		J.Virol.1981.40:551-559 & Eur.J.Biochem.1979.101:59-64 & Microbial Drug Resistance 1997.3:165-176
	HB-623 & HB-746		J.Virol.1990.64:5149-5155
	EJ-1**		J.Bacteriol.1992.174:5516-5525
	Dp-2 & Dp-4		J.Virol.1978.26:221-225
	Dp-1		Virology 1975.63:577-582
	w-3 & w-8		J.Virol.1976.19:659-667
	304		J.Bacteriol.1980.141:1298-1304
	HB-1, HB-2, HB-3**, HB-4, HB-5 & HB-6		J.Bacteriol.1979.138:618-624
<i>Streptococcus pyogenes</i>	T12**		Mol. Microbiology. 1997#23:719-728
	A-1	12202-B1	The American Type Culture Collection
	A-6	12203-B1	
	A-25	12204-B1	
<i>Streptococcus sp./Enterococcus</i>	Kjem	14918	
	1	HER 339	Felix d'Herelle Reference Centre, Quebec, Quebec
	182	HER 80	
	VD1884	HER 323	
	1A	12169-B1	The American Type Culture Collection
	1B	12170-B1	
	NN	21597-B1	
	42	19948-B1	
	118	19951-B2	
	120	19952-B1	
<i>Veillonella rodentium</i>	N2		Antonie Van Leeuwenhoek 1989.56:263-271
<i>Vibrio cholerae</i>	Psi 92		Intervirology 1993.36:237-244
	VCB-1,2,3 & 4		J.Infection 1998.36:131
	CP-T1**		J.Virol.1984.51:163-169
	VSK		FEMS Microbiol.Lett.1996.145:17-22
	Phi138		J.Virol.1986.57:960-967
	Phi149		J.Virol.1985.140:217-223
	Fs-2**		Microbiology 1998.144:1901-1906

	e4 e5 X29 β κ 13 14 16 24 32 57		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio cholerae</i> (Cont'd)	138 145 149 163 N-4 S-5 S-20 M-4 D-10 I II III IV V	14100-B1 14100-B2 14100-B30 14100-B4 51352-B1 51352-B2 51352-B3 51352-B4 51352-B5 51352-b6 51352-B7 51352-B8 51352-B9 51352-B10	The American Type Culture Collection
<i>Vibrio costicola</i>	UTAK		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio eltor</i>	e ₄		J.Gen.Virol.1987.68:1411-1416
<i>Vibrio natrigens</i>	nt1, nt6		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio parahaemolyticus</i>	KVP40** VF33 VP1 ϕ 60 ϕ HAWI-5 ϕ PEL8C-1		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio sp.</i>	α 3a		Felix d'Herelle Reference Centre, Quebec, Quebec
	NN	11985-B1	The American Type Culture Collection
	ph1	51582-B1	
	Phi149		J.Virol.1987.61:3999-4006
<i>Veillonella rodentium</i>	N2		Antonie V.Leeuwenhoek.1989.56:263-271

<i>Yersinia enterocolitica</i>	1		Felix d'Herelle Reference Centre, Quebec, Quebec
	2		
	3		
	4		
	5		
	6		
	7		
	8		
	9		
	φYeO3-12		
	I, IV & VIII		Zentralbl.Bakteriol.Mikrobiol.Hyg.1982.253:1 02
<i>Yersinia pestis</i>	R	23208-B1	The American Type Culture Collection
	S	11593-B1	
	Y	23053-B1	
	II		Zh.Mikrobiol.Epidemiol.Immunobiol.1990.11 :9
<i>Yersinia pseudotuberculosis</i>	PST**	23207-B1	The American Type Culture Collection
<i>Yersinia sp.</i>	RD2		Mol.Gen.Mikrobiol.Virusol.1990.8:18-21

xxxx)

Table 2

>Bacteriophage 77, complete genome sequence, 41708 nucleotides

```
1 gatcaaaata cttgggggag gggttagggag taaacttcgc gataatttta aaaattcatg
61 tataaccccc ctcttataac ctttttaagg cagggtgatga aatggagatt atagtcgatg
121 aaaatttagt gcttaaaaga aaagaaagac tacaagtatt atataaagac atacctagca
181 ataanttaaa agtagttgat ggtttaatta ttcaagcagc aaggctacgt gtaatgcttg
241 attacatgtg ggaagacata aaagaaaaag gtgattatga tttatttact caatctgaaa
301 eggcgccacc atatgaaagg gaaagaccag tagccaaact atttaagtct agagatgctg
361 caratcaaaa aataatcaaa caattatcgg atttattgcc cgaagagaaa gaagacacag
421 aaacgcacac tgatgattac ctatgattag taataaatac gttgatgaat atataaattt
481 gtggaaacaa ggaaagataa ctttaataaa agaaagaatt gatctcttta attatctaca
541 aaaaacatata tattcccgag atgatgtata ttttgatgaa cagaaaatcg aggatgttat
601 caaattttatt gaaaaatggc attttccaac attaccattt caaagggtta tcatagctaa
661 tatattttct atagataaaa atacagatga agctttcttt acagaatttg ctattttcat
721 gggacgtgga ggcggggaaa acgggtctaact aagtgctatt agtgattttc ttctacgccc
781 ctacacgga gttaaagaat atcacatctc cattgttgct aatagtgagg atcaagcaaa
841 aacatcgttt gatgaaatca gaaccgtttt aatggataac aaacgaata agacgggtta
901 aacgccaana gctccttatg aagttagtta agcaaaaaata ataaacgctg caactaaact
961 gggtattcga tataacacat caaacacaaa aaccaaagac ggtggacgtg aggggtgtgt
1021 tatttttgat gaaattcatt atttctttgg tccgaaatg gtaaacgtca aacgtggtgg
1081 attaggtaaa aagaaaaata gaagaacgtt ttatataagt actgatggtt ttggttagaga
1141 gggcttatat gatgcaatga agcacaanaat tgcaagtgtt ttaagtggca aggttaaaaa
1201 tagtagattg ttctgctttt attgttaagt agacgatcca aagaagttg atgacagaca
1261 gacgtgggaa aaggcgaaac caatgttaca taaacgttta tcagaatacg ctaaaacact
1321 gctaagcacg attgaagaag aatataacga ttaccatttc aaccgttcaa ataagcccca
1381 attcatgact aagcgaatga atttgcttga agttgccttt gaaaaagtaa tagcaccatg
1441 gaaagaanaa ctacgcacta atagagagat accaaattta gataatcaaa tgtgtattgg
1501 tgggttagac ttgcnaaca ttcgagattt tgcaagtgtt gggctattat tccgaaaaaa
1561 cgatgattac atttggttag gacattcggt tgaagaacaa ggggttttgg atgatgtcaa
1621 attagaacct cctattaaag aatgggaaaa aatgggatta ttgaccattg tcatgatga
1681 tgcatttgaa attgaatata agtttgattg gtttttaaa gctagagaaa aatatgggtt
1741 tgaanaagtc atagctgata attatagaac tgatattgta agacgtgcgt ttgaggatgc
1801 tggcataaaa cttgaagatc ttagaatccc aaagaacata catggattac ttgcaccacg
1861 tatcgataca atgttttgcga aacataacgt aatatatgga gacaatcctt tcatgctgtg
1921 gtttactaat aatgtttgctg taaaaatcaa gccggatgga aataaagagt atatcaaaaa
1981 agatgaagtc agcgcataaa cggatggatt catggctttt gttcacgcac tatatagagc
2041 agacgatata gtagacanaa acatgtctaa agcgttgat gcatatga gtatagattt
2101 ctaatagagg aggtgagaca tgagtattct aganaagata ttaaaaacta ggaaagatat
2161 aacatatatg cttgatttag atatgataga agatctatca caacaagcgt atgtgaacag
2221 tttagcgatt gatagttgta ttgaatttgt tgccgcagct gtcgctcaa gtcattttaa
2281 agtattggaa ggtaatagaa ttcaaaagaa tgatgtttac tacaagttaa atataaaacc
2341 aaatactgac ttatcaagcg atagtttttg gcaacaagtt atataaaac taatttatga
2401 taacgaggtt ttaatcgtag taagtgcag caaagaatta cttatcgag atagctttta
2461 cagagaagag tacgctttgt atgatgat atccaagat gtaacggtta aagattatc
2521 ttatcaacgt actttcaaaa tgcgaaggt catatattta aagtacaaca acaataaagt
2581 gacacacttt gtagaaagtc tattcgaaag ttacgggaaa atattcgga gaatgatagg
2641 tgcacaatta aaaaactatc aaataaaggg gattttgaaa tctgcctcta gcgcataatg
2701 cgaaaagaat atagaaaaat tacaagcgtt cacaaataaa ttattcaata ctttttaata
2761 aaatcaacta gcaatcgcg ctttgataga aggttttgat tatgaggaat tatctaattg
2821 tggtaagaat agtaacatgc ctttttctga attgagtgag ctaatcgag atgcaataaa
2881 aaatgtrtgc ttgatgattg gtatacctcc aggtttgatt tccggagaaa cagctgattt
2941 ggaaaaaaac acgcttgat ttgagaagtt cgttttaaca cctttatata aaagattcca
3001 gaacgaatta aacgcgaaac tcataacaca aagcatgtat ttgaaagata caagaataga
3061 aattgtcgtt gtgaataaaa aagaccact tcaatagct gaagcaattg acaaaactgt
3121 aagttctggt tcatttcaaa ggaatgaggt gcggtattat ttagggtgaa aacctcaga
3181 caatcctgaa tttagcgaat acctgattac taaaaactac gaaaaagcta acagtgggtg
3241 aaatgatgaa aaagaaaaag atgaaaaaac ttgaaaggt ggtgatgag atgaaagcgg
3301 agattaaagg cgtcatcgtt tccaacgaag ataaatgggt ttacgaaatg cttggtatgg
3361 attcgacttg tccataaagt gttttaaac aactagaatt tagtgatgaa gatgttgata
3421 ttataattaa ctcaaatggt ggttaacctag tagctggtag tgaaatatat acacatttaa
3481 gagctcataa aggcaagtg aatgttcgta tcacagcaat agcagcaagt gcggcatcgc
3541 ttatcgcaat ggctggtgac cacatcgaaa tgagtcgggt tgctagaatg atgattcaca
```

5

142

10

15

20

25

30

35

40

45

50

55

3601 atcccttcag tattgcgcaa ggagaagctga aagatctaaa tcatgctgca gaaacattag
3661 aacatgttgg tcaataaatg gctgaggcat atgctggttag agctggtana aacaaacaag
3721 aactttataga aatgatggct aaggaaacgt ggctaaatgc tgatgaagcc attgaacaag
3781 gcttttcgga tagtaaaatg ttgaaaaacg acaatatgca aattgtatga agcgatacac
3841 aagtgctatc gaagatgta taaatcgtg taacagcttt ggtaagtaaa acgccaagagg
3901 tcaacattga tattgacgca atagcaataa aagtaattga aaaaataaat atgaagaaa
3961 aggaaccaga aatcgatgtt gcagatagta aattatcagc aaatggattt tcaagattcc
4021 ttttttaata caaaaatagg aggtcataaa atgactataa atttatcgga aacattcgca
4081 aatgcgaaaa ocgaatttat taatgcagta aacaacggtg aaccgcaaga aagacaaaat
4141 gaattgtacg gtgacatgat taaccaacta ttgaagaaa cttaattaca agcaaaaagca
4201 gaagctgaaa gagtttctag ttacctaata tcagcacaaa ctttgagtgc aaaccaagaa
4261 aatttcttta tggatatcaa taagagtggt ggatataaag aagaaaaact tttaccagaa
4321 gaaacaattg atagaattct cgaagattta acaacgaatc atccattatt agctgactta
4381 ggtatttaaaa atgctgggtt gcgtttgaag tcttaaaat cogaactctc tggcgtgggt
4441 gtttggggta aatctatagg tgaatttaaa ggtcaattag atgctgcgtt cagtggaaga
4501 acagcaattc aaaaataatt gacagcggtt gttgttttac caaaagattt aatgatttt
4561 ggtcctcgct ggattgaaag atttgttcgt gttcaaatcg aagaagcatt tgcagtggcg
4621 cttgaactgt cggttctaaa aggtactggt aagaaccaac cgattgggtt aaaccgtcaa
4681 gtacaaaaag gtgtatcggt aactgatggt gctatccag agaaagaaga acaaggtacg
4741 cttacatttg ctaatccggt cgctacggtt aatgaattga cgaagtggtt taaatccac
4801 tcaactaacg agaaaggtaa atcagtagcg gttaaaggtt atgtaacaaat ggttgtaaat
4861 cgttcgagtg cttttgaggt tcaagcacag tacaacattt taaatgcaaa tggcgtatat
4921 gttactgctt taacttttaa ttgaatgtt attgagttc cagtccaaga agcaggttaag
4981 gttttaacgt acgttaaaag tctatatgat ggttatttag ctggtgggtt caatgttcag
5041 aaaaattaaa aaacacttgc gttagatgat atggatttat acactgcata acaatttgct
5101 tacggcaaa cgaagataa taaagtgtct gctgtttgga aattagattt aaaaaggcat
5161 aaaccagctt tagaagatac cgaagaaaca ctataaaatt ttatgaggtt ataaatgggt
5221 gaaatttaaa gttgttagag aatttaaaaga atcagagcac aatcaacaca agtacaagaat
5281 agggggagttg tatccagctg aagggtataa caatcctcgt gttgaattgt tgacaaatca
5341 aatcaaaaat aagtacgaca aagtttatat cgtacctta gataagctga caaaaacaaga
5401 attattagaa ctatcgcaat cattacaaaa aaaaagcgtc agttcaattg ttaaaagtga
5461 aatcatcgac ttattgaatg gtgaagacaa tgacgattga tgatttgctt gtcnaattta
5521 aatcacttga aagattgac cataattcag aggatgagta cttaaaagcag ttgttaaaaa
5581 tgcgtacga cgtataaaa aatcagtggt gaggttttga attagagaat ttaatagggt
5641 aagaattgat actataacg gctagatag ctatcaaga ttattagaa cacttcaacg
5701 acaattacag acctgaataa atagattttt cgttatctct aatggaggta tcagaagatg
5761 aagaagagtt ttaagaaacc tagaattaca actaaacgtt taaatagcgt tgttcatctt
5821 tataagtata ctgaaaaata tgggtccagaa gctggagaaa aagaagaaaa attattatat
5881 agctgttggg cgagtattga tgggtctcgt ttacgtgaat tagaacaagc tatctcaaac
5941 ggaacgcaaa atgacataa attgtatatt cgtgatccgc aaggtgatta ttaccacgt
6001 gaagaacatt atctgaaat tgaatcaaga tattcaaaa atcgtttgaa tataaagcaa
6061 gtatcaccag atttgataa taaagacttt attatgattc gggagggata tagttcatga
6121 gtgtgaaggt gacaggtgat aaagcattag aaagagaatt agaaaaacat ttggcataa
6181 aagagatggt aaaaagttcaa gataaggcgt taatagctgg tgctaaggta attgttgaag
6241 aataaaaaaa acaactcaaa ctttcagaag actcaggagc actgattagt gaggattggc
6301 gtactgaacc tgaatggata aaggggaaac gtactgttac aattaggtgg cgtgggcctt
6361 ttgaacgatt tagaattagta catttaattg aaaaaggtca tgttgagaaa aagtcaggaa
6421 aatttgtaaa acctaaagct atgggtggga ttaatagagc aataagacaa gggcaaaaata
6481 agtatatttga gacgctaaaa agggagttga aaaaattgtg attgatattt tgcacaaagt
6541 tcatgaagtg actagtcag acagaatttat tagagagcac gtaaatatca ataattataa
6601 gttcaataaa tacctaantg taaaagatac tgatgtacct ttatttgta ttgacgatat
6661 cgacgaccca atacctacaa ctctacttga cggagatgag tgcacataa gttatattgt
6721 ccaaatagat gtttttgta agtacaatga tgaatataat gcgagaatca taagaataaa
6781 gatattctat cgcattcaaa agttattatg gtctgaacta aaaaaggaaa atgtttcaaa
6841 tggaaaaaccg gaatatatag aagaatttaa aacatataga agctctcggt ttacagagg
6901 cattttttat aaggaggaaa attaaatggc agtaaaacat gcaagtgctgc caaaggcgt
6961 tattaacatt actggtttag gtttcgctaa attaacgaaa gaaggcgcgg aattaaaaat
7021 tagtgatatt acaaaaaaaa gaggattaca aaaaattggt gttgaacctg gttggagaact
7081 aaaaacagct tatgctgatg gcggtccaat tgaatcaggg aatcacagcg gagaaggtaa
7141 aatctcatta caaatgcgat cgttccctaa agagattcgc aaaaattggt ttaatgaaga
7201 ttatgatgaa gatggcggtt acgaagagaa acaaggtaaa caaaacaatt acgtagctgt
7261 atgggttcaga caagagcgta aagacggtac atttagaaca gttttattac cttaaagttat
7321 gtttacaat cctaaaacg atggagaaac ggctgagaaa gattgggact tctcaagtga
7381 agaggttgaa ggtgaggcac ttttcccttt agttgataat aaaaagtcag tacgttaagt
7441 tatctttgat tcagctaaca tgacaaatca tgatggagac ggtgaaaaag gcgaagaggc
7501 tttcttaaa aaaaatttag gcgaagata tactggaaac gtgacagagg gtaacgaaga
7561 aactttgtaa caaaaccggc ttcatcgaa actgcggtaa agtcggttaa tataccagat
7621 agcattaaaa cacttaaaat tggcgacaca tacgatttaa atgttgatg agagccatct

5

10

15

20

25

30

35

40

45

50

55

7681 aatcaaagta agttattgaa atacacacga gatcaaacga atattgtatc aatcaatagt
7741 gatgggtcaag ttactgcgga agcacacaggc attgctacgg ttaaagcaac agtgggtaat
7801 atgagtgaca ctataacaaat aaatgtagaa gcataagagg gggcaacccc tctattttat
7861 ttgaaaataa ggagagtagt ataaaatggc aaaaataaaa cgtaacatta tcaaatagat
7921 agaagatcca aaagcaaatg aaattaaatt acaaacgtac ttaacaccac acttcatttc
7981 atttgaattt gtatacgaag caatggattt aatcgatgat attgaggacg aaaaatagcac
8041 gatgaagcca agagaatcgc ctgacagatt gatggatag gtgttaaaaa ttacagataa
8101 ccaattcaca gttaaagacc taaaagaacg tatgcatgca cctgatggaa tgaatgcact
8161 tctggaacaa gtgattttca ttactcaagg tcaacaaact gaggaaacta gaaattttat
8221 ccagaacatg aaaaagacc tgaagattta acatataaag caatgttgaa aaatcggat
8281 actctcatga tggacttaat tgaataatgg aaagacgcta acgaagtgtt aaaaatgccca
8341 tttcattatg tgcctttccat atatacaaat aaaaataatg acatttctga agaaaaagca
8401 gaggtcttaa ttgatgcatt ttaaccttaa ccgtttgggt aggggtatttt tttgaacct
8461 ttttagaaaag gaggtaaaaa atggggagaaa gaataaaagg tttatctata ggtttggatt
8521 tagatgcagc aaatttaaat agatcatttg cagaatacaa acgaaacttt aaaaacttaa
8581 attctgactt aaaaataaca ggcaacaact tcaaatatac cgaaaaacta actgatagtt
8641 acaaaacaaag gattaaagaa cttgatggaa ctatcacagg ttaataagaa aacgttgatg
8701 atttagccaa gcaatatgac aaggtatctc aagaacaggc cgaaacagct gcagaagctc
8761 aaaaagttac scaagaatat aacaaacaaag caaatgagct gaattattta gaaagagaat
8821 tacaataaac atcagccgaa tttgaagagt tcaaaaaagc tcaagttgaa gctcaagaa
8881 tggcagaagag tggctgggga aaaaaccagta aagtttttga aagtatggga cctaaattaa
8941 caaaaaatggg tgatgggtta aaatccattg gtaaaaggtt gatgtatgg gtaactgcac
9001 ctgttttagg tatgacagca gcatacaggaa aagcttttgc agaatgttat aaggttttag
9061 atactgttac tcaagcaaca ggccgaacag gcagtgaatt aaaaaaattg cagaactcat
9121 ttaagatagt ttatggcaat ttccagcag atgctgaact tgttggtgga gttttaggag
9181 aagttaatc aaggttaggt ttacacagta aagaacttga aatgccaca gagtcatctt
9241 tgaattcag tcaataaca ggttctgagc gtgtgcaagc cgtacagta attaccgtg
9301 caatgggcga tgcaggtatc gaagcaagtg aatatcaag tgctttggat atggtagcaa
9361 aagcggcgca agctagtggt ataatgttg atacattagc tgatagtatt actaaatcag
9421 gcgctccaat gagagctatg ggctttgaga tgaagaatc aattgcttta tctctcaat
9481 gggaaaaatc aggcgttaat actgaaatag cattcagtg tttgaaaaaa gctatatcaa
9541 attggggtaa agctggtaaa aacccaagag aagaatttaa gaagacatta gcagaattg
9601 aaaaagcgcg gatatagct agcgaacaa gtttagcgt tgaagcattt ggtgcaagg
9661 caggtcctga ttagcagac gctattaaag gtggtcgtt tagttatcaa gaatttttaa
9721 aaactattga agattcccaa ggcacagtaa accaaacatt taaagattct gaaagtggct
9781 ccgaaagatt taaagttaga atgaataaat taaaattagt aggtgctgat gtatgggctt
9841 ctattgaaag tgcgtttgct ccgtaaatg aagaattaat caaaaagcta tctatagcgg
9901 ttgatttggt ttccaaattta agttaggtt ctaaaagatc aattgttatt ttcagtggta
9961 ttgctgtcgc aattggctcc gtatgttttg ggttaggtgc atttataagt acaattggca
10021 atgcagtaac tgtattagct ccattgttag ctagtattgc aagggtcgtt ggaattgata
10081 gttttttatc gactaaagta cctataatag gaactgtctt cacagcttta actggctcaa
10141 ttggcattgt attaggtgta ttggctgggt tagcagtcgc atttacaatt gctataaga
10201 aaactgaaac attttagaat ttgtttaatg gtgcaattga aagtgttaaa caaacattta
10261 gtaattttat tcaattttat caacctttcg ttgattctgt taaaaacatc ttaaaacaa
10321 cgatatcagc aatagttagt ttgcgaaaag atatttggag tcaaatcaat ggattcttta
10381 atgaaaacgg aatttccatt gttcaagcac ttcaaaatc atgcaacttt attaaagcga
10441 catttgaatt tatttttaaat ttgttaatta aaccaattat gttcgcgatt tggcaagtga
10501 tgcaatttat ttggccggcg gttaaagcct tgattgtcag tacttgggag aacataaaag
10561 gtgtaataca aggtgcttta aatatcatac ttggcttgat taagtctctc tcaagcttat
10621 tctgttggtg ttggcgagga gtttgggacg ccgttggtgat gattctttaa ggaagcagttc
10681 aatttaattg gaatttagtt caattatggt ttgtaggtta aatacttggt gctgttaggt
10741 actttggcgg gttgctaaaa ggattgatag caggaaattt ggacgttaata agaagtatat
10801 tcaagtaaac ttatcagca atttggatg caacaaaaag tatttttggg tttttattta
10861 atagcgtaaa atcaattttc acaaatatga aaaaattggt atctaatact tggagcagta
10921 tccgtacgaa tacaatagga aaagcgcagt cattatttag tggcgtcaaa tcaaaattta
10981 ctaatttatg gaatgcgacg aaagaatttt ttagtaattt aagaatttgg atgtcaata
11041 tttggaattc cattaaagat aatacggtag gaattgcaag ccgtttatg agtaaggatc
11101 gtggaatttt cacaatatg cgcgatggct tgaattccat tatagataag attaaaagtc
11161 atatcggcgg tatggttaag gctattaaaa aaggacttaa taaattaatc gacggtttta
11221 actgggtcgg tggtaagttg ggaattgata aaatacctaa gttacacact ggtacagagc
11281 acacacatac tactacaaga ttagttaaga acggttaagat tgcacgtgac acattcgcta
11341 cagttgggga taagggaacg ggaattgggt caaatgggtt tagaatgaa atgattgaat
11401 tccctaacgg taaacgtgta atcacaccta atacagatc taccgcttat ttacctaaag
11461 gttcaaaagt atacaacggt gcacaaactt attcaatggt aaacggaaacg cttccaagat
11521 ctatgttagg tactatgtgg aaagatatta aatctgggtc atcatcggca tttaactgga
11581 caaaagataa aataggtaaa ggtacaaatt ggcttggcga taaagtggc gatgttttag
11641 attttatgga aaatccaggc aaacttttaa attatatact tgaagctttt ggaattgatt
11701 tcaattcttt aactaaaggt atgggaattg caggcgacat aacaaaagct gcattggtcta

5

10

15

20

25

30

35

40

45

50

11761 agattaagaa aagtgtact gattggataa aagaaaattt agaagctatg ggcggtggcg
11821 atttagtcgg cggaaatatta gacctgaca aaattaatta tcattatgga cgtaccgcag
11881 cttataccgc tgcacttggg agaccatttc atgaagggtg cgtatttcca ttgtatatc
11941 aagaagttag aacgcgatg ggtggcagac ttacaagaat gccatttatg cctggtgggt
12001 atggttaatta tctaaaaatt actagtggcg ttatcgatat gctatttcgc catttgaaaa
12061 actttagcaa atcaccacct agtggcacga tggtaaaagcc cgggtgatgt gttggtttaa
12121 ctggtantac cggatttagt acaggaccac atttacattt tgaatgagg agaaatggac
12181 gacattttga ccttgaacca tatttaagga atgctaagaa aaaaaggaga ttatcaatag
12241 gttgtggcgg tgctacttct ggaagtggcg caacttatgc cagtcgagta atccgacaa
12301 cgcaaatgat tttaggtggg cgttataaag gtaaatggat tcatgaccaa atgctgcgcg
12361 ttgcacaacg tgaagtaaac taccagtcga atgcagtga taactgggat ataaatgctc
12421 aaagaggaga cccatcaaga ggtattatcc aaatcatcgg ctcaactttt agagcaaacg
12481 ctaaaagcgtg atatactaac tttaataatc cagtacatca aggtatctca gcaatgcagt
12541 acattgttag acgatattgt tgggtgggtt ttaaacgtgc tgggtattac gcatatgcta
12601 caggtggaaa agtttttgat ggttgggtata acctagggtga agacgggtcat ccagaaatgga
12661 ttattccaac agatccagct cgtagaatg atgcaatgaa gattttgcat tatgcagcag
12721 cagaagtaag agggaaaaaa gcgagtaaaa ataaagcgtcc tagccaatta tcagacttaa
12781 acgggtttga tgactcttag ttattattga aaatgattga acaacagcaa caacaatag
12841 cttctattact gaaaatagca caatctaacg atgtgattgc agataaagat tatcagccga
12901 ttattgacga atacgtcttt gataaaaagg tgaacgcgtc tatagaaaag cgaagaaggc
12961 aagaatcaac aaaaagtaag tttagaaaag gaggaaatgc ttttcaatga tagacactat
13021 taagtgtaac acaaaaaaaa ttcccttgggt gcatgtcgaa agaggggttg aaataccctc
13081 ttctaattat gtttataaaa cagaataatg agatggagct tgggggtcta tatataaagg
13141 gcgtaggctt gaattctata gttttgatat acctttggtg gtacgtaatg actatttctc
13201 tcacaacggc attaaaacac atgatgacgt cttgaatgaa ttagtaaaatg tttttaacta
13261 cgaggaacaa gttaaattac aattcaaatc taagattgg tactggaacg cttatttcga
13321 aggaaccaata aagctgcaca aagaatttac aaatccgtgt aagttcacta tcaaatagat
13381 actaacagac ccttacaat attcagtaac aggaataaaa aatactgcga tttcagacca
13441 agtttcagtt gtaaatagtg ggaactgcta cactccctta attgttgaa cccgagcaat
13501 taacacatct agttacttta tgattactaa aaatgatgaa gattatttta tggttggtga
13561 tgataggata accaaagaag ttaaggatta catgctcctt gtttatcata gtgagttctg
13621 tgatttcaaa ggttggacta agatgattac tgaagatatt ccaagtaatg acttagtggt
13681 taaggtcggc ggtgactttg tgatatccaa tcttggcgaa ggatataaag caactaat
13741 tcttgatgca aaaggttggg ttggtgctgg cagcaaaacga ggggtcccta aagcgatgac
13801 agattttcaa attacctata aatgtattgt tgaacaaaaa ggttaagggt cgggaagaac
13861 agcacaacat atttatgata gtgattgtaa gttacttctt tctattgggt atgaaaaata
13921 atacatgat agaaaaatag gacatattgt tgtacgttg tataaccaa agggagaccc
13981 caaaaagata tgcactatc agaataaacc gataatgtat aacttggaca gaactcgtgt
14041 ttatatgcgg ctcagaagag taggtaataa attttctatt aaaacttggg aatttgatca
14101 cattaaagac ccagatagac gtaaacctat tgatattggat gagaagaagt ggaatagatg
14161 cggtaagttt tatcagcgtc cagcttctat catagctgtc tatagtgcga agtataacgg
14221 ttataagtggt atggagatga atgggttagg ttcatcnaat acggagattc taccgaaaac
14281 gaaaagtcga agggatgta ttatacaaaa aggtgattta gtaaaaatag atatgcaagc
14341 aaaaagtgtt gtcacatgat aggaaccaat gttgagcgag aaatcgtttg gaagtaatta
14401 ttcaatggtt gattctgggt acagtgaaat aatcatacaa cctgaaaacg tctttgatac
14461 gacggttaaa tggcaagata gatatttata gaaaggagat gagagtgtga tacatgtttt
14521 agattttaac gacaagatta tagatttctt ttctactgat gaccttctct tagttagagc
14581 gattcataaa cgtaatgtta atgacaattc agaaatgctt gaactgtctc tatcatcaga
14641 aagagctgaa aagttccgtg aacgacatcg tgtattata agggattcaa acaaaacatg
14701 gcgtgaattt attattaact ggtttcaaga tacgatggac ggctacacag agatagaatg
14761 tatagcgtct tatcttgctg atataacaac agctaaaaccg tatgcaccag gcaaatttga
14821 gaaaaagaca acttcagaag cattgaaga tgtgttgagc gatcacaggt ggggaagttc
14881 tgaacaaacc gaatacgtat gcttacgtac tacgtcatgg acttcttctc aaactagata
14941 tgaagtttta aagcaattat gtacaaccta taaaatgggt ttagattttt atattgagct
15001 tagctctaact accgtcaaa gtagatatgt agtactcaaa aagaaaaaca gcttattcaa
15061 aggttaagaa atgaaatag gtaagattt agtcgggtta actaggaaga ttgatatgtc
15121 agaaatcaaa acagacttaa ttgctgtggg acctgaaaat gacaaaagga agcgtttaga
15181 gctagtgtgt acagatgacg aagcgcaaa gcaattcaac ctacctatgc gctatatgtg
15241 ggggatatat gaaccacaat cagatgatca aaatagtaat gaaacacgat taagtctctt
15301 agccaaaaca gagttaata aacgttaagtc ggcagttatg tcatatgaga ttacttctac
15361 tgattttgaa gttacgtatc cgcacgagat tatatcaatt ggcatgacag tcagagtaaa
15421 acatagagat ttttaaccgc cattgtatgt agaggcagaa gttattgctg aagaatataa
15481 cataatttca gaaatagca catatacatt cgggtcaacct aaagagttca aagaatcaga
15541 attacagaaa ggtttaaca agcagattga cataatacat caaaagttaa acgataatat
15601 tagcaatatt aacactatag ttaagatgct ttagatgggt gaattagaat actttgaacg
15661 caaaatacac aaagtgata caccgccaga aaatccagtc aatgatatgc tttggtatga
15721 tacaagtaac cctgatgtg ctgtcttgcg tagatatggg aatggcagat ggaattgaagc
15781 aacaccaaat gatgtgaaa aattaggtgg tataacaaga gagaagcgc tattcagta

55

5

145

10

15

20

25

30

35

40

45

50

15841 attaaacaat atttttatta attttatctat acaaacacgct agtcttttgt cagaagctac
15901 agaattactg aatagcgagc acttagtaga taatgatttg aaagcggact tacaagcaag
15961 ttttagacgt gtgattgatg ttataatca aattaaaaat aatttagaat ctatgacacc
16021 cgaactgca acgattgggc ggttggtaga tacacaagct ttatttcttg agtatagaaa
16081 gaaattacaa gatgtttata cagatgtaga agatgtcaaa atcgccattt cagatagatt
16141 taaattatta cagtcacaat acactgatga aaaaataaaa gaagcgttgg aaataatagc
16201 acaaaaattt ggtttaacgg tgaatgaaga ttgacagtta gtcggagaac ctaagtgtgt
16261 taaatcagct attgaaagc cttagagaatc cacaagaaga caattacgtg actatgtaaa
16321 aacatcggac tataaaaacag acaaaagcgg tattgttgaa cgtttagata ctgctgaagc
16381 tgagagaacg acttttaaaag gtgaatcaa agataaagtt acgttaaacg aatatcgaaa
16441 cggatttgaa gaacaaaaac aatatactga tgaccagtta agtgatttgt ccaataatcc
16501 tgagattaaa gcaagatttg aacaagcaaa tcaagaagcg caagaagctt taaaatcata
16561 cattgatgct caagatgatc ttaaaagaaa ggaatcgcaa cgttatgctg atggttaaat
16621 ttcggaagaa gagcaacgcg ctatacaaga tgctcaagct aaacttgaag aggcataaaa
16681 aaacgcagaa ctaaaaggcta gaaacgctga aaagaagact aatgcttata cagacaacaa
16741 ggtcaagaaa agcacagatg cacagaggaa aacattgact cgttatggtt ctcaaatat
16801 acaaaatggt aaggaaatca aattaagaac tactaaagaa gagtttaag caaccaatcg
16861 tacactttca aatataatga acgagattgt tcaaaatggt acagatggaa caacaatcag
16921 atactgatgt aacggagtggt ctcaagcttt gaattgtggg ccacgtggta tttagattaaa
16981 tctgtataaa attgatatta acggttaatag agaaataaac ctcttattcc aaatatgcy
17041 agataaagta gataaaacgg atattgtcaa cagctctta tttcaagag agggctcttg
17101 tatcaactgt aatagaattg gaattaaagg cggcgacaat aacgatatg ttcaaatata
17161 gaatgattct attgaactag gtggtattgt gcaacgtact tggagagga aacgttcaac
17221 agacgatatt tttacgcgac tgaaagacgg tcacctaaga tttagaata acaccgctgg
17281 cgggttcact tataatgtcac attttggtat ttcgacttat attgatggg aaggtgaaga
17341 cgggtggttca tctggtacga ttcaatggtg ggaataaaact tacagtgata gtggcatgaa
17401 tggataaaca atcaattcct atggtggtgt cgttcacta acgtcagata ataactgggt
17461 tgttctggag tcttacgctt catcgaaat caaaagcaaa caggccaccg tgtattata
17521 tccaaacaca gacaaagtg ctggattaaa ccgatttgca ttcacgctgt ctaatgcaga
17581 taatgcttat tccagtgacg gttatattat gtttgggttc gatgagaact atgattacgg
17641 tgcgggtatc aggtttttcta aagaaagaaa taaaggctct gttcaaatg ttaattggagc
17701 atatgcacaa ggtggagata caacaatcga agcaggggat ggcaaattta atatgctgaa
17761 acgacgtgat ggttaattag atattcatat acagagtaca gacctactgt ctgtagggtc
17821 agatgatgca ggaatagga tagcttctaa ctcaatttat agacgtactt attcggccgc
17881 agctaatttg catattactt ctgctggcac aattgggcgt tccacatcag cgcgtaata
17941 caagtattct atcgaaaatc aatataacga tagagatgaa caactgggaa attcaaaagc
18001 tattcttaac ttacttatta gaacgtggtt tgataaagct gagtctgaaa ttttagctag
18061 agagctgaga gaagatagaa aattatcgga agcacctat aaacttgata gatagctagg
18121 ttgtattgtc gaagagggtg agaatttagg attaaaagag ttgtcactgt atgatgaaa
18181 agggagaatt gaaggtatag cgtatgatcg tctatggatt catcttacc ctgttatcaa
18241 agaaacaaca ctaagaatca agaaattgga ggaagtcaag aatgcaggat acaacaacag
18301 gattacaagc taatcctgaa tatacaattc attatttatc acaggaaatt atgaggttaa
18361 cacaagaaaa cgcgatgta aaagcgata tacaagaaaa taaagaaaa caacaatggt
18421 ctgaggagaa gtaatcctta gcactatttt tatacaadna ttaaggagg tcattraatt
18481 atggcaaaag aaattatcaa caatacagaa aggtttattt tagtacaatt cgacaagaaa
18541 ggtacagaac gtgtagtata tcaagatttc acagggaagt ttacaacttc tgaatgggtt
18601 aaccatgctc aagattttata atctgaagaa aacgctaaga aaattcgga gacgttaaat
18661 ttgttatatc aattaaactaa caaaaacaaa cgtgtgaaag tagtcaaga agtagttgaa
18721 agatcagatt tatctccaga ggtaacagtt aacactgaaa cagtatgaaa agctatgagt
18781 tagatactca tagtctttat tcttttagaa agcgggtgta ctgaattggg gtggttcaaa
18841 aaacacgaac atgaatggcg catcagaagg ttagaagaga atgataaaac aatgctcagc
18901 acactcaacg aaattaaatt aggtcaaaaa acccaagagc aagttaaat taaattagat
18961 aaacacctag atgctattca aaaagaaga gaaatagatg aaagaataaa gaaagaaaat
19021 gataagaaca tacytgatag gaaaatgtgg gtgcttggtt tagtgggac aatatttggg
19081 tgcctaatta tagcattatt gcgtatgctt atgggcataa aagaaggggt attaccatgt
19141 tcgattaaaa ttttgagctc tgcgttgga cgtgtttctg gtttggtaag tgtaagtaat
19201 agttaagagt cagtgtctgc gcactggctt tttattttgg ataaaaggag caaacaaaatg
19261 gatgcaaaag taataacaag atacatcgta ttgatcttag cattagtaaa tcaattctta
19321 gcgaacaaag gtattagccc aattccagta gacgatgaaa ctatatcatc aataatcctt
19381 actgtatgct ctttatatac aacgtataaa gacaatccaa catctcaaga aggttaaatg
19441 gcaaatcaaa aattaaagaa atataaagct gaaaataagt atgaaaagc aacagggcaa
19501 cgcgcaatta aagaagtaat gacacctacg aatatgaacg acacaaatga tttagggtag
19561 gtgggtgata tatgttaagt acaaaaaatc aagcagaaaa atggtttgac aattcattag
19621 ggaacaatc caacccagat ggttggtatg gatctcagtt ttatgattac gccaatatgt
19681 tctttatggt agcgacaggg gaaaggctgc aaggtttata tgcttataat atccggttg
19741 ataataaagc aagattgaa aatatgtgtc aataaataa aaactatgac agctttttac
19801 cgcnaaagtc ggaattgttc gttttccctc caaagtatgg tggcggagct ggaacggtg
19861 aaattgttga gagcgcaaat ttaaatatct tcacatcatt tggtaaaaa cggaaacggta

55

5

10

15

20

25

30

35

40

45

50

55

19921 aaggttggac taatggcggt ggcgaacctg gttgggggtcc tgaactctg acaagacatg
19981 ttcattatta tgacaatcca atgtatttta ttagggtlaaa ctccctaac aacttaagcg
20041 ttggcaataa agctaaaggt atctattaagc aagcgactac aaaaaagag gcagtaatta
20101 aaactaaaaa aattatgctt gtagccggtc atgggtataa cgtacctgga gcagtaggaa
20161 acggaacaaa cgaacgcgat ttatatacgt aatataaac gcctaatac gccaaagtatt
20221 taagacatgc aggaatgaa gttgcattat acggtgggtc aagcactca caagatatgt
20281 atcaagatgc tgcatacggg gttatgttag gcaataaaaa agattatggc ttatattggg
20341 ttaaatcaca ggggtatgac attgttctag aaatacattt agacgcagca ggagaagcg
20401 caagtgggtg gcatgttatt atctcaagtc aattcaatgc agatactatt gatataagta
20461 tacaagatgt tattaataat aacttaggac aaataagagg tgtgacctt cgttaagtatt
20521 tactaaatgt taatgtatca gcgaaataaa atataaatta tctgttattt gaattagggt
20581 ttattactaa taaaatgat atggattgga ttaaaaaaaa ctatgacttg tattctaat
20641 taatagccgg tgcgattcat ggttagccta taggtgggtt ggtagctggg aatgttaaaa
20701 catcagctaa aaaaaaaa aatccaccag tgccagcagg ttatacactc gataagaata
20761 atgtccctta taaaaaagaa caaggcaatt acacagtagc taatgttaaa ggttaataatg
20821 taagagacgg ttatccaact aattcaagaa ttacaggggg attaccacaa aacacaacaa
20881 ttacgtatga cgggtcatat tgtattaatg gttatagatg gattacttat attgttaata
20941 gtgggcaaac tctgtatata gcgacaggag aggtagacaa ggcaggtaat agaataagta
21001 gttttggtaa gtttagcagc atttagtatt tacttagaat aaaaattttg ctacattaat
21061 tactagggaat ctacagttta ttaataaact atttgatgg agtataatat tctatacac
21121 tttttaaact ttctctcaag atttaaatgt agataacagg caggactctc ggtacttggc
21181 tattttttta tcttatagct agccttccggg ctagtttttt gttatgatgt gttacacatg
21241 catcaactat ttacacttat cctgttctac ccaagcatgt cactggatgt ttttcttgc
21301 gatagagagc atagttttca tactactccc cgtagtatat atgactttag cattcccgta
21361 taacagttta cggggtgctt ttatgttata attgctttta tatagttaga tgaactata
21421 tagccgggca gaggccatgt atctgactgt tgggtccaca ggagacatct tcttctgtat
21481 cactcgatgc atatatctta acaacataga aatgttacct tgcctataac cgtatcttaa
21541 tgcgacgggt tatatttatt cccctacaac caacaaaacc acagatccta ttaatttagg
21601 atctgtggtta ttttttgggt ttttttgggg caaaaaaagg gcagattatt tgaaaaaagg
21661 caaacgcttg tggaaaagct aaaaagttta aatgacaaa aaccttgata caacagtgtt
21721 ttgggacgct cgtgtacgtt agagaatgac cgggtttacca tcatacaagg gtgggattaa
21781 cttgtgttaa aaagccttta atacagttg ttacaaagga tttgtagcgt ttttaaaat
21841 aaaaaagggc agaaaaaggc cagatacctt ttatgacaca agtttttcta atttttgcct
21901 taactctctg tccattttct ctgttacctg tgtatacacc ttatagtcg tttttctac
21961 tgtatgtctt actcttttca taattgtctt taacgatata ttcatttccg ccaataaact
22021 tatgtgtgta tgccttagtg tgtgagtagt aactttttta tttatattta atgattctgc
22081 agctgaggac aatcgtttgt ttactctact gccttgata ggaatttctt ggcaagtgtt
22141 gaataaatac cctctatcaa catagcttgg ttcccatgtt tgcatttttt tattttctaa
22201 cattattttt ttcaatacat ttgctatctt tgaattgatg gcgatttttc tcttgaacc
22261 tgcggttcta gtagtatctt tgtgacaaa tccagcatta catttgatc tgtgaatagt
22321 gccatttaata gcgacgtttt tatttttgag gtcaacatct ttaacttgga gagtaataa
22381 ctacactatg cgcatacctg ttaaagctcg aacttttaca gccccagcaa cttaaaatcag
22441 agctctatac tgcattgtat tatcgttcag tataaaatcg cgtatctgta ttacctgttc
22501 catctctaaa tagttatata ttttcgcttc ttttttttct atactctcta tctcttact
22561 ctcttttggg agtgtgacgc tatttaatat gtgttctgtt ggataattgt aaaaattaac
22621 gggtatttta atagcttctt tcatatgtcc aagttgacgc tttaactgat tgcagaata
22681 caogtttgat aattgtttaa taaatgtttg catgtacttt gtatcaattt tgtttaaag
22741 taaattttga gaactgttct ttttgatgtt ttgattcttt gttttcaa tacaagcgt
22801 cgttacttta aagccagatg tttttatatg atattcaagc cattcatcta ataacgcgtg
22861 aaaaagcaaa gtttttaatt cgtttgacga cttgttgggt agtttttctt ttatttttct
22921 ttctaaacga aacattgcct ctttttgcga ttgctttgta ttcttatcca agacaacact
22981 tacacgtttc catttatctg tatagggatc ttgtatttc tctgtatctc tatactctgt
23041 ttcatgttct ttatttttaa atttttcaaa ccacatttta catccctcct caaaattggc
23101 aaaaaataat aagggttagc gggctaccca tgaataatgt ataaaaaag acgctgtat
23161 aaaaacaga cgcacttat aattataaga ttacatgggt aattaccaaa atgggtaacg
23221 aatatatacg tgttttaaag gataaacctt taatatatta aaattatctc atcttatctc
23281 agggatctgc aatatattat tattaattct atttatcagt aacataatat ccgaagaatc
23341 tattactgga tttttaattt ttgggggtaa aacttttctt atcgaaaact tactaatcgg
23401 ctggaaagaa tttatgcaag cgttaactatt accctttaat ttttttacct tatcaattgc
23461 tgatactatg ttattaatgt ttctgtcaat ttattttaat ttattttcaa ttcttaact
23521 atcagatata aattcaataa aataactctt agtgatgaat tctgtgtgtt ttttttggta
23581 ttttttatcg aaaaactctt ttaatatagc tgaattattt tgcgcgctaa ttaattttaa
23641 aaacactctt aaataactct cccatttcaa atcaaaatc atctttaaat acttttgtt
23701 ttcttttagg gataaggaa taacatttac tatatcttcc gttatagaat catttttatt
23761 catcactatt gcaagtggt aattagaaaa ttctttatta acgtttatcc cgaactctac
23821 aaaaactatt tctctctgtt taaactttgg ataaaaact ttatgggttt tttcaccttc
23881 aaactctctg agtaaatagt gaactctga atctaacttt ttaaattttg gattccaga
23941 agtttttaac ttattaatgc gtttttctat attacgcgtc atcatttctc ctctattctc

5

147

10

15

20

25

30

35

40

45

50

24001 gctcacactc tcaccaccat tcaacgtcta cacttgtagg cgttttttga ttagtaaaat
24061 cataatgaat cttcttttgg taaacttatcg ccatctattt ttgtgaaat aaattccaag
24121 tattttacgag catttatgtga cgataaaatct ttaggttaact cataagttaa tgggttgatta
24181 ccactagttta aaacttcata tactatagtt tcttttttta ttttgcaact agttattttc
24241 attataaact ccttttaaac actgctgaaa tagacgtctt tttcaaataa gcatgattaa
24301 tacttttaatt ctttaatcca catatattta aaagttaggt agtaggtaac aaataaaga
24361 ctttaagttta agattgcttt ttcatgtca atttctctt tgtttatatt tatattaaag
24421 cgctaaatat acgtttactaa tcacaataca accttgccca ttactttaat atcactaaac
24481 gaagcgactt tgatatcatc atacttcgga tttagagata ccaaattaat atagtcttcg
24541 catatatcta cagcgttgat aagacttact ccatcctaata caacgagtcg aattgtacca
24601 tcttttaacag aatcttcttt ctttaaaaaa gcgtatgttc cttgttttaa cataggttcc
24661 attgaatcac cattaaactaa aatacaaaaa tcagcatctg atggcgcttc gtcttcttta
24721 aaaaataact cttcatgcaa tatgtcatca tataattctt cctctatgcc agcaccagtt
24781 gcaaccacatg caatatacga tactagttra gactctttat attcatctat agaagtgaat
24841 ttactcttgt catctaatg cccatttgca tagttaagta cgttttcttg gcggggaggt
24901 gtgagtttag aaaaatgtt attgattttt gacattatcg ttcatcttg acgttcttcg
24961 tcaggaaactc gataagaatc tacatcatc cccataagcc acgcttcacc gacattctaa
25021 gttttatgata ataagaataa tttatgttgg tctggagaag accttcatt aacatactgg
25081 gataagtgac tttttgacat ttttaatttc aattcttttt gaaaggggtt cgaactttct
25141 agaataatcta cttgacgcaa gttcctatct ttcatattt gttttaatct ttcaagaagt
25201 ttttgcattg gtaatgcctc cttgaatttc attatatagg aagggaaata aaaaatcaata
25261 caaaagtcca accttttttaa ctttttgtgt tgacatttgt caaaattggg gttatagtta
25321 ttatagtcca aatgtttgaa cttaggaggt gattatttga atactaatac aacttttgat
25381 ttttctgtat tgaacggtaa gatagtcgaa gtgtactcga cacaatttaa ctttgcata
25441 gctttagggt tctcagaaag aactttgtct ttgaagttga acaacaaagt acctgggaaa
25501 acaacagaca ttattaaagc ttgttaagtt ttgggaatac ctataaaga tgttcacaaa
25561 tattttttta aacagaaagt tcaaatgttt gaacttaata agtaaggag gcataacaca
25621 tgcaagaacy agaaaaggtt aataaaagta acacatcttc aaatgaagca tcaaaacctt
25681 ttaggacaaa ttgaagctta cgacaaaacg ctttaagaaa taagttacac tcgagacctt
25741 tacaacaaac acctaaagcat gaacaaacga gacgatttcg ctggtttgga aatggtagag
25801 gatgaaatta ctaaaaagct acgaagtgtc atcaaaagtt tccaaaaagt agtgaagcgc
25861 tttagacaagc ttaacgggtg tgaagcgat aacaaagtta ctgatttaac agagtggcgc
25921 aaagtgaatc agtaacatct accttctaat ataaccacgc ttatcaacat ccacattgag
25981 cagatgtgag cgaagagctgg cgatgatag agccgcgttt aaatacattc gatagtcatt
26041 gcgataaacg tctgctgaat gtgggtgttg aggaanaag aggatactca aatgcaagca
26101 ttacaacat ttaattttta agagctacca gtaagaacag tagaaattga aacgaaacct
26161 tattttgtag gaaaagatat tgcgtgatt ttaggatag caagatcaaa caatgcatt
26221 agaaatcatg ttgatagcga ggacaagctg acgcaccaat ttatgtcatc aggtcaaaac
26281 agaaatcatg tcattatcaa cgaatcagga ttatcacagtc taattcttga tgcctctaaa
26341 caaagcaaaa acgaaaaaat tagagaaacc gctagaaat tcaaacgcgt ggtaacatca
26401 gatgtctac cagctatttc caaacacggt atatacgcaa cagacaatgt aattgaacaa
26461 acattaaaag atccagacta catcattaca gtgttgactg agtataagaa agaaaaagag
26521 caaaacttac ttttcaaca gcaagtagaa gttcaacaa caaaagtatt attcgctgac
26581 tcggtagctg gtagtgataa ttcaatactt gttggagaac tagcgaatat acctaaacaa
26641 aacggtgttg atataggaca aaacagattg ttcaaatggt taagaaataa tggatatctc
26701 attaaaaaga gtggagaaaag ttataactta ccaactcaaa agagtatgga tctaaaaatc
26761 ttgatataca aaaaacgaat aattaataat ccagatggtt caagttaagt atcacgtaca
26821 ccaaaaagtaa caggcaaaag acaacaatac ttgttaata agtttttagg agaaaaacaa
26881 acatcttaaa aggaagaaac caatggaaca aatcacatta accaaagaag agttgaaaga
26941 aattatagca aaagaagtt gagaggtat aaatggcaag aaaccaatca gttcaggttc
27001 aactttcagt aaagttaagaa tcaataatga cgatttagaa gaaatcaata aaaaactcaa
27061 tttcgcaaaa gatttgtcgc taggaagatt gaggaaagctc aatcatcga ttcgctaaa
27121 aaagtatcag catggcttcg aatcaattca tcaaaaagct tatgtacaag atgttcata
27181 ccatattaga aabtaacat tatcaatttt tggagtgaac cttaattcag acttgagtga
27241 aagtgaatac aacctagcag caaaagtta tcgagaatc aaaaactatt atttatacat
27301 ctatgaaaag agagtttcag aattaactat cgatgatttc gaataaagga ggaacaacaa
27361 atgttcaaaa aatttagaat tgcgaagaa aaaaaataat taaaactcaa attactcaag
27421 catgctagt actgttaga aagaaacaac aaccctgaac tgttgcgagc agttgcagag
27481 ttgtgaaaa aggttagcta aattcaacggt taaggttttg ccttgctcct acacttagag
27541 ttgagatcc aacaaacaca taagtttttag tagggcttag aaaaaatgtt tcatcttctt
27601 cttttgtaac agtttcaatt cttcatatc ctggaaaaac aattttcttt aaatccgaaa
27661 catgtttttt tgaacccatc ttttaagtaa ctagaagttt catactatc acctccttag
27721 gttgataaca acattataca cgaagggagc ataaacata tgcaagcatt acaaaacaa
27781 tcgaacatcg gagaatgtt caatattcaa gaaaaagaaa atggagaat cgcaatcagc
27841 ggtcgagaac ttcatcaagc attagaagtt aagacagcat ataaagatt gtttccaaga
27901 atgcttaaat acggaattga agaaaaata gattacacag ctatcgctca aaaaagagca
27961 acagctcaag gcaatagac tcaactatg gaccacgac tcacactaga cactgcaaaa
28021 gaaatcgcaa tgattcaac tagtgaaact ggcaaacgt caagacaata tttcatccaa

55

5

10

15

20

25

30

35

40

45

50

28081 gttgaaaaag catggaacag cccagaaatg attatgcaac gtgctttaaa aattgctaac
28141 aacacaatca atcaattaga aacaaagatt gcacgtgaca aacaaaaaat tgtatttgca
28201 gatgcagtag ctactactaa gacatcaatt ttagttggag agttagcaaa gatcattaaa
28261 caaaacggta taaacatcgg gcaacgcaga ttgtttgagc gggtacgtca aaacggattc
28321 ctattataac gcaaggggtg ggattataac atgcctacac agtattcaat ggaacgtgag
28381 ttattcgaaa ttaagaaac atcaatcaca cattcggacg gtcacacatc aattagtaag
28441 acgccaanaag taacaggtaa aggacaacaa tactttgtta acaagttttt aggagaaaaa
28501 caaacaactt aataggagga attacaatg aacgcactat acaaaaaaac cctccrctac
28561 acaatggcag ttgtgacgtg gaaggtttgg aagattgaga agcacactag aaaaacctgtg
28621 attagtagca gggcgttgag tgactatcta aacaacaaat ctttaacctat accgaagatg
28681 gctgaaaaat ctactgaatc tgctcgtcgc cttttgaagt tcgcccgaaca aactattagc
28741 aaatacaaac attatacacg aaaggaaaga tagaaatgcc aaatatcata gtaccaccaa
28801 caccagaaaa cacatataga ggcgaagaaa aattttgtga aaagtatac gcaacaccta
28861 cacaatcca tcaattgttt ggagtatgta gaagtacagt atacaactgg ttgaatattt
28921 accgcaaga taatttgggt gtgaaaaatt tatcatttga ttattcaccac acaggcactc
28981 tgatttaaat ttctaaattg gaagagtatt tgatcagaaa gcataaaaaa ttgtattagg
29041 aggatattaa atgagcaaca ttataaaa ctacctagta gcagtattat gcttcacagt
29101 cttagcagatt gtacttatgc cgtttctata cttcactaca gcattgtcaa ttgctggatt
29161 cgcagtagtc gcaacattca tgactacaaa agaattgcttt ttcaagaat aaaaaaactg
29221 ctacttgggt gagcaagtaa cagtatcaaa cacttaagaa aaaaattcatg ttcaataaa
29281 aacgaaaaac ggaaggagtc aagatgtatt acgaaatagg cgaatcata cgaaaaata
29341 ttcatgttaa cggattcgat tttaagctat tcattttaaa aggtcatatg ggcatatcaa
29401 tacaagttaa agatatgaac aacgtaccaa ttaaacatgc ttatgtcga gatgagaatg
29461 acttagatat ggcattcagac ttatttaacc aagcaataga tgaattggatt gaagagaaca
29521 cagacgaaca ggacagacta attaacttag tcatgaaatg taggaggtgc gctatgaagc
29581 agactgtaac ttatatcatt cgtcataggg atatgccaat ttataaact acaaaccaa
29641 ctgatacaaa ttcatgattt agttactcca caaatagaaa tagagctagg gagtttaacg
29701 gtatggaaga agcagtagtc aataggtatt atcacaagac aatcaagaaa acagtgaacg
29761 aaactattga gtacgaggag gtgaaacatg actgaggaac aacaagaacc acaagaaaaa
29821 gtaagcatat tcaaaaaact aaagataaat aatctcgtcg agaaaaataa aaggaaattc
29881 tacaattttg cagtatacgg aaaaattggc tcaggaaaaa ccacgtttgc tacaagagat
29941 aaagacgctt tcgtcattga cattaacgaa ggttgaaaca cggttactga cgaagatca
30001 gaogtagaaa tcgagaacta tcaacacttt gtttatgttg taaatttttt acctcaatt
30061 ttacaggaga tgagagaaaa cggacaagaa atcaatgttg tagttattga aactattcaa
30121 aaacttagag atatgacatt gaatgatgtg atgaaaaata agtctaaaaa accaacgttt
30181 aatgattggg gagaagttgc tgaacgaatt gtcagtatgt acagattaat aggaaaactt
30241 caagaagaat acaaatccca ctttgttatt acaggtcatg aaggtatcaa caagatataa
30301 gatgatgaag gtgacactat caacctact atcactattg aagcgcaaga acaaatataa
30361 aaagctatta cttctcaag tgatgtgtta gctagggcaa tgattgaaga attgatgat
30421 aacggagaaa agaaagctag atattttcta aacgtgaac cttctaatac gtttgaaaca
30481 aagattagac attcaccttc aataacaatt aacaataaga aatttgcaaa tcctagcatt
30541 acgagcgtag tagaagcaat tagaaatgga aactaaaaat taattaaaag gacggatttt
30601 aattatgaaa atcacaggac aagcgcaatt tactaaagaa acaaatcaag aaaaagttta
30661 taacggctca gcagggtttc aagctggaga attcacagt aaagttaaaa atattgaatt
30721 caatgataga gaaaaatag atttcacaat cgtatttgaa aatgatgaag gcaaacata
30781 taaacataat caatttgtac cgcgtataa atatgatttc caagaaaaac aattgattga
30841 attagttact cgaattaggta ttaagttaaa tcttcctagc ttagattttg ataccaatga
30901 tcttattggg aagttttgtc acttggattt gaaatggaaa ttcaatgaag atgaaggtaa
30961 gtattttacg gatttttcat ttattaaacc ttacaaaaag ggcgatgatg ttgttaacaa
31021 acctattccg aagacagata agcaaaaaag tgaagaaaat aacggggcac aacaacaaac
31081 atcaatgtct caacaaagca atccatttga aagcagtggc caatttggat atgacgacca
31141 agatttagcg ttttaagggtg tggtttaaat gcaatcactt acaagatacc agaaagataa
31201 cgacggtagt tattccgtcg ttgctactgg tgttgaaact gaacaaagtc acattgactt
31261 actagaaaaa ggaatccac taaaagcaga agtagagggt ccggacaata aaaaactatc
31321 tatagaacaa cgcaaaaaaa tattcgcaat gcttagagat atagaacttc actggggcga
31381 accagtagaa tcaactagaa aattattaca aacagaattg gaaattatga aaggttatga
31441 agaattcagt ctgocgagct gttctatgaa agttgcaagg gagttaatag aactgattat
31501 agcgtttatg ttctcatcctc aaatacctat gagtgtagaa acgagtaagt tgttaagcga
31561 agataaagcg ttattatatt ggggtacaaat caacgcgaac tgtgtaatat gcggaagccc
31621 tcacgcagac ctggcacatt atgaagcagt cggcagaggg atgaacagaa acaaaatgaa
31681 ccactatgac aaacatgtat tagcgtttatg tcgcgaacat cacaacgagc aacatgcgat
31741 tggcgttaag tcgtttgatg ataaatacca cttgcacgac tcgtggataa aagttgatga
31801 ggggtctaat aaaaatttga aaggagagaa aaaggaatga atagactaag aataataaaa
31861 atagcactcc taactcgtat cttggcggaa gagattagaa atgctatgca tgcgttaaaa
31921 gtggagaaaa ttttaaaatc tccgttttagt taatacaggt ttttaaaaaa gctttaccat
31981 aggcggacaa actaatttag ccttttttga tgcctattac ccaggggctg taatgttaact
32041 ttaatacttc aaattcaatg ccagaaagtt tacttattgt tcttaggttg tgcctgact
32101 ttaacattct tttacaaat tctaattccc aaacaaatct ttgtttttct ataattttat

55

5

149

10

15

20

25

30

35

40

45

50

32161 taaagtgtatt taaaaactga ggagcataaa acctattata aattcccttt tttgttaagt
32221 aagacatgttc aaaagtttta tttaaaaccc ctaaccttac taggttatta attgaaattt
32281 cggtttgattc tatatctaac ggagagttct ttattaacgt gtccgataata ttcataccgt
32341 cattctttgg gtttaaaacc gccctatatt taacggcagg atgtacttcg tgattcttta
32401 aatgttttaa aagaatagca tcatttgggg ataattgttt aattatttca acaaatgaat
32461 ggtgggttaa tgagtttttt ctgtcatcca tagatgatgc tattagtttt gcgaacatat
32521 tacttaaaagt tttttcacta atgtaaaact ttgaagcttc tagagcaggc cctagaagag
32581 aaaattgtgg tttttgtaaa ttatttttag gtacagaaga tattttcttt tcaaatgttt
32641 ctttgaattt ttcaaattct acttctcttt gataataaac tttatccaca taaaggttga
32701 atttcccaaa gacaagttcc caagtttttag agaattgttc tacaggccct tttgatgcgc
32761 ctccaataat tttatcaata cctttaccta aaataggatc cataattatt caccaccaat
32821 ctaacgcaat agcgataata aaattatacc agaaaggaga atcaacatga ctgaccaacc
32881 aagttactac tcaataatta cagcaaatgt cagatacagat aaccgactta ctgacagcga
32941 aagttactt tttgcagaaa taacatcttt aagtaacaaa tacggatact gcacagcaag
33001 taatggttac tttgcaactt tatacaacgt tgrtaaggaa actatatctc gtgaattttc
33061 gaacctttacc aactttgggtt atctaaaat cgaaattatc aaagaaggtta atgaagttaa
33121 acaaaaggaa atgtaccctt tgacgcaaac gtcaataact attgacgcaa aaatcaatac
33181 ccctattgat aattctgtca atacccttat tgacgcaaat gtcaagaga atattacaag
33241 tattaataat acaagtaata acaatataaa tagaatagat atattgtcgg gcaacccgac
33301 agcatcttct ataccctata aagaatttat cgattactta acaaaaaaag cgggcaagca
33361 ttttaaacac aatacagcta aaacaaaaga ttttattaaa gcaagatgga atcaagattt
33421 taggttggag gattttaaaa aggtgattga tatcaaaaca gctgagtggt taaacacgga
33481 tagcgataaa tactcttagac cagaaaacact tttggcagat aattttgagg ggtacctcaa
33541 tcaaaaaata caaccaactg gcacggatca attggaacgc atgaagtacg acgaaagtta
33601 ttgggattag ggggatatta tgaacacact attcagcgaa aagataaacg aaagcttgaa
33661 aaaaatacaa cctactcatg tcgaaaaagg attgaaatgt gagagatgtg gaagtgaata
33721 cgacttatat aagtttgctc ctactaaaaa acaccggaat ggttacgagt ataaagacgg
33781 ttgcaaatgt gaaatctatg aggaatataa gcgaacaaag caaccggaaga taaacaacat
33841 attcaatcaa tcaaacgtta atccgtcttt aagagatgca acagcctaaa actacaagcc
33901 acaaaatgaa aaacaagtac acgctaaaaca aacagcaata gagtacgtac aaggcttctc
33961 tacaanaaga ccaaaatcat taattattga aggttcatat ggaacttgga aaagccacct
34021 agcatacgtc atcgcaaaag cagtcaaaagc taaagggcat acggttgctt ttatgcacat
34081 accaattgtg atggatcgta tcaaaagcac atacaacaaa aatgcagtag agactacaga
34141 cgagctagtc agattgctaa gtgatattga tttacttgta ctatagata tgggtgtaga
34201 aaacacagag cacactttaa ataaactttt cagcattgtt gataacagag taggttaaaa
34261 caacatcttt acaactaact ttagtataa aagaactaat caaaataga actggcaacg
34321 tataaattcg agaatgaaaa aaagagcaag aaagtaaga gtaactcggag acgatttcag
34381 gggagcgagat gcatggtaac caaagaattt taaaaaacta aacttgagtg ttcagatag
34441 tacgctcaga aactcataga tgaggcagag ggcgatgaaa ataggttgta cgacctattt
34501 atccaaaaac ttgcagaacg tcatacacgc cccgctatcg tcgaatatta aggagtgtta
34561 aaaaatgccg aagaaaaata ttactttatc cgagaagatg gcacagaaga tattaaggtc
34621 atcaagtata aagacaacgt aaatgaggtt tattcgctca caggagccca tttcagcgac
34681 gaaaagaaaa ttatgactga tagtgacctt aaacgattca aaggcgctca cgggcttcta
34741 tatgagcaag aattaggttt acaagcaacg atatttgata tttagagggt gacgatgagt
34801 aaatacaacg ctaagaagaat tgagtacaaa ggaattgtat ttgatagcaa agtagaggt
34861 gaattattac aatattttaga aagtaatatg aatggcacta attatgatca tatcgaaata
34921 caacogaat tcgaattatt accaaaacta gataaacac gaaagattga atatatgca
34981 gacttcgctg tatatctcga tggcaaacgt attgaaagtt tcgacattaa aggtatgcca
35041 accgaagtag caaaacttaa agctaaagat ttcagacata aatacagaaa caaaaaactc
35101 aatttgatgt gtaaaagcgc taagtataca ggtaaaaat ggattacgta caggaataa
35161 attaaagcaa gacgagaacg caaaaagaaa atgaagtgtat ctaatgcaac aaacagcata
35221 tataaatgca acgattgata taaggatacc tacagaagtt gaatatcagc attttgatga
35281 tgtggataaa gaanaagaag cgctggcaga ttacttatat acaaatccrg acgaaatact
35341 agagtatgac aatttaaaaa ttagaacgt aaatgtagag gtggaataaa cgggcagtggt
35401 tgcaatcatt aataataaac catataaatt taacaatttt gaaaaaagaa ataatggcaa
35461 agcgtgggat aaatgctgga attgtttcta aacgtgttag aggtgtgtgg gaggtttcag
35521 aagcttttaga cggccttat ggcattgcacc taaaagaata tagagaaatg aaacaaatgg
35581 aaaaatttaa acaagcgaga ctcgaaactg aattggaaaag agagcgaaag aaagaggctg
35641 agctacgtaa gaagaagcca cattttgtta atgtacctca aaacatttca cgtgatccgt
35701 actggttcga tgtcacttat aaccaaatgt tcaagaatat gagtgaagca taatgagcat
35761 aatcagtaac agaaaagtag atatgaacaa aacgcaagac aacgttaagc aacctgcgc
35821 ttacacatac ggcgacattg aaattataga ttttattgaa caagtacgg cacagtaccc
35881 accacaatta gcaattcgaa taggtatagc aattaaatc ttgtctagag caccgttaaa
35941 gaatggtcat gaggatttag caaaggcga gttttacgtc gatagagtat ttgacttgt
36001 ggagtgatga ccatgacaga tagcggacgt aaagaatact taaaacottt tttcggctct
36061 aagagataac tgtatcagga taacgaacga gtggcacata tccatgtagt aaatggcact
36121 tattactttc acggtcatat cgtgccaggt tggcaaggtg tgaanaagac atttgataca
36181 gcggaagagc ttgaacata tataaagcaa agtgatttgg aatatgagga acagaagcaa

55

5

150

10

15

20

25

30

35

40

45

50

36241 ctaactttat tttaaaaggg cggaaacaat gaaaatcaaa attgaaaaag aaatgaattt
36301 acctgaactt atccaatggg cttgggataa ccccaagtta tcaggtaata aaagattcta
36361 ttcaaatgat gttgagcgca actgttttgt gacttttcat gttgatagca tcttatgtaa
36421 tgtgactgga tatgtatcaa ttaacgataa atttactgtt caagaggaga tataacaatg
36481 aaaatcaaa ttaaaaaaga aatgagatta gatgaattaa ttaaatgggg gcgagaaaaat
36541 cgggatctat cacaaggaaa aatatttttt tcaacaggat ttagrtgagg attcgttcgt
36601 tttcatccaa atacaataaa gtgttcgacg tcaagtttta ttccaattga tatccctctc
36661 atagtgtgata ttgaaaaaga agtaacggaa gagactaagg ttgatagggt gattgaatta
36721 ttcgagattc aagaaggaga ctataactct acactatatg agaactactag tataaaagaa
36781 tgtttatatg gcagatggtt gcctaccaaa gcattctaca tcttaaacga tgacctaaat
36841 atgacgttaa tctggaaaaga tggggagtgt ctagtatgat gttgaatttt aaagcttggg
36901 ataaagataa aaaagtattg agtattattg acgaaatcga ttttaaatgt gggatcattt
36961 tgaatttcaa aggttataaa agttcaatg aagtaaaact attacaatc acaggattta
37021 aagatgtgca cgggtcggag atttatgaag gggatattgt tcaagattgt tattcgagag
37081 aagtaagttt tatcgagttt aaagaaggag ccttttatat aacttttagc aatgttaactg
37141 aattactaag tgaaaatgac gatattattg aaattgttgg aaattatttt gaaaaatgaga
37201 tgctattgga ggttatgaga tgacgttcac ctatcagat gaacataata aaaatccttg
37261 tactaactct aacaagttat tagataaact tcacaaagca ttaaaagatc gtgaagagta
37321 caagaagcaa cggagatgagc ttattgggga tatagcgaag ttacgagatt gtaacaaaga
37381 tctagagaag aaagcaagcg catgggtag gtattgcaag agcgtgaaa aagatttaat
37441 aaacgaattc ggttaacgat atgaagagat taaattcggg atggaattaa acaataaaat
37501 ttttatggag gatgacacaa atgaataatc cggaaaaaat cgaacagtcg gttatttagt
37561 ctagtgcgta taacggtaat gacacagagg ggttgctaaa agagattgag gacgtgata
37621 agaagcgca agcgtttgat gaaatacttg agggaaatgac aaatgctatt caacattcag
37681 ttaaaagagg tattgaactt gatgaagcag tagggattat ggcaggtcaa gttgctata
37741 aatatgagga ggaatagaaa aatgactaac acattacaag taaactattt atcaaaaaat
37801 gctagatgc ccgaacgaaa tcataagacg gatgcagggt atgacatatt ctcagctgaa
37861 actgtcgtac tcgaaccaca agaaaaagca gtgatcaaaa cagatgtagc tgtgagtata
37921 ccagagggct atgtcggact attaaactagt cgtagtgggt taagtagtaa aacgtattta
37981 gtgattgaaa caggcaagat agacgcggga tatcatggca atttagggat taatatcaag
38041 aatgatgaag aacgtgatgg aatacccttt ttatatgatg atatagacgc tgaattagaa
38101 gatggattaa taagcatttt agatataaaa ggttaactatg tacaagatgg aagaggcata
38161 agaagagttt accaaatcaa caaaggcgat aaactagctc aattgggtat cgtgcctata
38221 tggacccggg aactaaagca agtggaggaa ttcgaaagtg ttccagaacg tggagcaaaa
38281 ggcctcggaa gtagcggagt gtaaaagcat cttagatcga gttaaaggag ttttggggaa
38341 gtgacgcaat acttactcac aacattcaaa gattcaacag gacgaccaca tgaactattt
38401 actgtggcta gagataatca gacgtttaca gttattgagg cagagagtaa agaagaagcg
38461 aaagagaagt acgagggaca agttaaaaga gatgcagtta ttaaaagggt tcagttgat
38521 gaaaaataaa gggagtgtgg gaaatgacg atgttaaaat taaaactatt tcaggtggag
38581 tttattttgt aaaaacagct gaaccttttg aaaaatatgt tgaagaatg acgagtttta
38641 atggttatat ttacgcaagt actataatca agaaaccaac gtattataaa acagatacga
38701 ttgaatcaat cacacttatt gaggagcatg ggaatgaat cagctgagaa ttttattaca
38761 tgacggtagt agtttgatat tacatgaaga tgaattattt aacgaatatg tatttgtttt
38821 ggacaatttt agaaatgatg atgactattt aacgatagaa aaagattatg gcagagaact
38881 tgtattgaac aaaggttata tagttgggat caatgttgag gaggcagatg atgattaaac
38941 tacctaanaa gaaattcccg aaaaagtaca ctgaaataat caaaaaatat aaaaaaaag
39001 caccctgaaga aaaggtcaag attgaagatg attttattaa agaaattaaa gataaagaca
39061 gtgaatttta cagtcctacg atggctaata tgaatgaata tgaattaaag gctatgttaa
39121 gaatgatgcc tagtttaatt gatactggag atgacaatga tgattcaaaa accttaaaat
39181 atggatgggt tcgacatctt tattgttga atactgtcat tattcgggtt atttcgattg
39241 ctacttgcta tcacattgcc tatctataca gtggctagtt accaacaaca agaattacat
39301 caaggaacta ttacagataa atataacag agacaagata aagaagacaa gttctatatt
39361 gtattagaca acaaacagat cattgaaaat tccgacttat tattcaaaaa gaaatttgat
39421 agcgagata tacaagctag gttaaaagta ggcgataagg tagaagttaa acaaatcgg
39481 tatagaatc acttttttaa tttatatccg gtcttatagc aagtaaaaga ggtagataaa
39541 caatgattaa acaataacta agaccattat tcttactagc aatgtatgag ttaggtaagt
39601 atgtaactga gcaagtgtat attatgatga cggctaata tgaatgagag gcgcccagtg
39661 attacgtctt tcgagcggag gtgagtgaat aatgagaata tttatttatg atttgatcgt
39721 tttgctgttt gctttcttaa tatccatata tattattgat gatggagtga taataaatgc
39781 attaggaatt tttggtagt ataaaattat agattccttt tcagaaaaa ttataaagag
39841 gtagataaaa atgaacgagc aaataacagg aagcatatat acttttagcag gaggtgttgt
39901 gctttattca gttaaagaga tttttaggta ttttacagat tctaacttac aacgtaaaaa
39961 aatcaattta gaacaaatat atccgataata tttagattgt tttaaaaagg ctanaaagat
40021 gattggagct tatattatc caacagacac gcatgaattt ttagattttt ttgatattga
40081 agtctttaat aatttagata agcaaaagtaa aaaaagcgtat gaaatgtta ttggatttag
40141 acaaatgatt aatttatcaa atagagttaa ggcattggaa gattttaaga tgagtttcaa
40201 caatgaattt agtacaaatc agattttttt taatccttct tttgttatgg aaacaattgc
40261 tattataaat gaatatcaaa aagatatatc ttatttaaaa aatataatta ataaaaatga

55

5

151

10

15

20

25

30

35

40

45

50

55

```
40321 tgaanaataga gcttataatc atattgatac ttttatcact tcagagtcacc gacgaaaaat
40381 aaacgattat aatctttatc ctgataaatt tgaagaacag tttagtcaaa agttttaaatt
40441 aaacagaaact tcgataaaag aaagaattat tattaattta aacaagagga gattttaaatt
40501 atgtggatta ctatgactat tgtatttgct atattgctat tagtttgat cagtattaat
40561 agtgatcggt caagagagat acaagcactt agatatacga atgattatct acttgatgaa
40621 gtagttaaaa ctaaaagggtt caacgggtta gaagaataca ggattgaatt gaagcgaatg
40681 aataacgata ttaaaaaagta atttatatta tcggaggtat tgcattgaat gataaagatt
40741 gagaacacag atatacaaaa gcttgaagaa cacattcagc acatcgataa ctatcgaaag
40801 gagttgaaga tgcgagaata tgaattactt gaaagtcag aaccagataa tgcgggagct
40861 ggcaaaaagta atttgccggg taaccogatt gaacgatgtg caataaagaa gtttagtgat
40921 aacaggtaata atacattaag aaatatagtt aacgggtgtg atagattgat aggtgaaagt
40981 gatgaggata cgcttgagtt attaagggtt agatattggg attgtcctat tgggtgttat
41041 gaatgggaag atatagcaca ttactttggt acaagtaaga caagtatatt acgtagaagg
41101 aatgcactga tcgataagtt agcaaaagta attgggtatg ttagcggac ttttacccta
41161 tgtaagtcag cattaataca gtttattatg ttagtatcag attaatattt aaagtattta
41221 aatgctaata cgacgcata acaagaggcg catcactatg ttagtgtctt ttttatttat
41281 gaggtatgaa catgttcaaa ctaattgtta atacattact acacatcaag tatagatgag
41341 tcttgatact acttaagtta tataagggtg aacattatga tgactaaaga cgaacgtata
41401 cgattctata agtctaagaa atggcaata acaagaaaaa gagtgttaga aagagataat
41461 tatgaatgtc aacaatgtaa gagagacggc aagtttaaga catatgacaa aagcaagcgt
41521 aagtcgttgg atgtagatca tatattatcg ctagaacatc atccggagtt tgctcatgac
41581 ttaaacatt tagaaacact gtgtattaaa tgtcacaca aaaaagaaa gagatttata
41641 aaaaagaaa ataaatggaa agacgaaaaa tggtaaatat ccccggttca aaaaatcaa
41701 aagcgatc
```

Table 3

5

10

15

20

25

30

35

40

45

50

55

Name	Position	Name	Position
1 77ORF005	19572..21026	48 77ORF052	1762..2013
2 77ORF006	3976..5196	49 77ORF053	37521..37757
3 77ORF007	21871..23076	50 77ORF054	22818..23060
4 77ORF008	2120..3307	51 77ORF055	17546..17788
5 77ORF009	31946..32803	52 77ORF058	18892..19122
6 77ORF010	26092..26889	53 77ORF059	34564..34785
7 77ORF011	24441..25208	54 77ORF064	29574..29795
8 77ORF012	29788..30576	55 77ORF065	28528..28746
9 77ORF013	33620..34399	56 77ORF066	27494..27703
10 77ORF014	27760..28512	57 77ORF069	38341..38547
11 77ORF015	3291..4028	58 77ORF070	36269..36475
12 77ORF016	32867..33610	59 77ORF071	40498..40701
13 77ORF017	23269..23982	60 77ORF072	38735..38938
14 77ORF018	31169..31840	61 77ORF073	30945..31148
15 77ORF019	39851..40501	62 77ORF074	38544..38738
16 77ORF020	6926..7570	63 77ORF075	13673..13870
17 77ORF021	37762..38304	64 77ORF077	25357..25605
18 77ORF022	30605..31156	65 77ORF079	29089..29280
19 77ORF023	26903..27346	66 77ORF080	35204..35389
20 77ORF024	10700..11140	67 77ORF085	24060..24242
21 77ORF025	9707..10147	68 77ORF092	39706..39876
22 77ORF026	40729..41145	69 77ORF094	32226..32393
23 77ORF027	6518..6925	70 77ORF096	13606..13773
24 77ORF028	34795..35199	71 77ORF098	7092..7256
25 77ORF029	6117..6521	72 77ORF102	29051..29212
26 77ORF030	36478..36879	73 77ORF104	34393..34551
27 77ORF031	39151..39546	74 77ORF109	18282..18434
28 77ORF032	33892..34266	75 77ORF112	39543..39692
29 77ORF033	5758..6120	76 77ORF117	27361..27501
30 77ORF034	7886..8236	77 77ORF118	38390..38530
31 77ORF035	19258..19560	78 77ORF120	36059..36199
32 77ORF036	36876..37223	79 77ORF124	33699..33833
33 77ORF037	102..446	80 77ORF128	14221..14355
34 77ORF038	34908..35219	81 77ORF130	15675..15806
35 77ORF039	37220..37528	82 77ORF133	8414..8542
36 77ORF040	41377..41676	83 77ORF140	13113..13235
37 77ORF041	35454..35753	84 77ORF147	7029..7148
38 77ORF042	5490..5774	85 77ORF149	30668..30787
39 77ORF043	29304..29564	86 77ORF151	31837..31953
40 77ORF044	18481..18768	87 77ORF155	30278..30391
41 77ORF045	5216..5500	88 77ORF157	4044..4157
42 77ORF046	25663..25935	89 77ORF167	20692..20799
43 77ORF047	11159..11425	90 77ORF175	35717..35821
44 77ORF048	28776..29039	91 77ORF176	6836..6940
45 77ORF049	36013..36255	92 77ORF178	35390..35491
46 77ORF050	35753..36007	93 77ORF179	8318..8419
47 77ORF051	38931..39167	94 77ORF182	29268..29564

5

153

Table 4

10

77ORF017 sequence

15

20

25

30

35

40

45

50

55

```
23982      atgacgcataatatagaaaaacgcattaataaattaaaaacttct
1   M   T   H   N   I   E   K   R   I   N   K   L   K   T   S
23937      ggaaatccaaaattttaaaaagtttagattcagatattcactattta
16  G   N   P   K   F   K   K   L   D   S   D   I   H   Y   L
23892      ctcaagagatttgaaggtgaaaaaaaccataaagggtttttatcca
31  L   K   R   F   E   G   E   K   N   H   K   G   F   Y   P
23847      aagtttaacaaggagaaatagttttttagatttcggtataaac
46  K   F   K   Q   G   E   I   V   F   V   D   F   G   I   N
23802      gttaataaagaatttttctaattcacactttgcaatagtgtgaat
61  V   N   K   E   F   S   N   S   H   F   A   I   V   M   N
23757      aaaaatgatttctaatacggaggatatagtaaaatgttattccctta
76  K   N   D   S   N   T   E   D   I   V   N   V   I   P   L
23712      tcctctaaagaaaaacaaaaagttttaaagatgaattttgatttg
91  S   S   K   E   N   K   K   Y   L   K   M   N   F   D   L
23667      aaatgggagttatttttaagattgttttttaatttaattagcgcg
106 K   W   E   Y   Y   L   R   L   F   L   N   L   I   S   A
23622      caaaataattcagctatatttaaagaagttttcgataaaaaaac
121 Q   N   N   S   A   I   L   K   E   V   F   D   K   K   Y
23577      caaaaaaacaacacagaattcatcactaaagattattttattgaa
136 Q   K   N   N   T   E   F   I   T   K   D   Y   F   I   E
23532      tttatatctgtagtttagaaattgaaaataaattaaataaaatt
151 F   I   S   D   S   L   E   I   E   N   K   L   N   K   I
23487      gacagaaacattaataacatagtatcagcaattgataaggtaaaa
166 D   R   N   I   N   N   I   V   S   A   I   D   K   V   K
23442      aaattaaaaggtaatagttacgcttgcataaattctttccagccg
181 K   L   K   G   N   S   Y   A   C   I   N   S   F   Q   P
23397      attagtaagtttcgcataagaaaagttttacccccaaaaattaaa
196 I   S   K   F   R   I   R   K   V   L   P   Q   K   I   K
23352      aatccagtaatagattcttcggatattatgttactgataaataga
211 N   P   V   I   D   S   S   D   I   M   L   L   I   N   R
23307      attaataataatatattgcagatccctgatataagatga 23269
226 I   N   N   N   I   L   Q   I   P   D   I   R   *
```

Physico-chemical parameters of ORF 77ORF017

1 MTHNIEKRIN KLKTSGNPKF KILDSDIHYL LKRFEGEKNH KGFYPKPKQG EIVFVDFGIN
 61 VNKEPSNSHF AIVMNKNDNH TEDIYVVIPL SSKENKKYLK MNFDLKWEYY LRLFLNLISA
 121 QNNSAILKEV FDKKYQKNNT EPITKDYFIE FIDSLSLEIN KLNKIDRNIN NIVSAIDKVK
 181 KLKGNYSYACI NSFQPISEKFR IRKVLPOKIK NPVIDSSDIM LLINRINNNI LQIPDIR

Number of amino acids: 237
 Average molecular weight (Daltons): 27887.38
 Mean amino acid weight (Daltons): 117.67
 Monoisotopic molecular weight (Daltons): 27869.83
 Mean amino acid monoisotopic weight (Daltons): 117.59

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	5	2.11%	7.58%	Cys	C	1	0.42%	1.66%
Asp	D	14	5.91%	5.28%	Glu	E	13	5.49%	6.37%
Phe	F	16	6.75%	4.09%	Gly	G	6	2.53%	6.84%
His	H	4	1.69%	2.24%	Ile	I	29	12.24%	5.81%
Lys	K	33	13.92%	5.95%	Leu	L	19	8.02%	9.42%
Met	M	4	1.69%	2.37%	Asn	N	30	12.66%	4.45%
Pro	P	7	2.95%	4.9%	Gln	Q	6	2.53%	3.97%
Arg	R	8	3.38%	5.16%	Ser	S	17	7.17%	7.12%
Thr	T	5	2.11%	5.67%	Val	V	11	4.64%	6.58%
Trp	W	1	0.42%	1.23%	Tyr	Y	8	3.38%	3.18%

Number of acidic (negative) amino acids (ED): 27
 11.39%
 Number of basic (positive) amino acids (KR): 41
 17.30%
 Total charge (KRED): 68
 28.69%
 Net charge (KR - ED): 14
 5.91%
 Theoretical pI: 10.01
 Total linear charge density: 0.30
 Average hydrophobicity: -5.37
 Ratio of hydrophilicity to hydrophobicity: 1.41
 Percentage of hydrophilic amino acid: 57.81%
 Percentage of hydrophobic amino acid: 42.19%
 Ratio of %hydrophilic to %hydrophobic: 1.37

5

155

77ORF019 sequence

39851 atgaacgagcaataataggaagcatatatacttttagcaggaggt
10 1 M N E Q I I G S I Y T L A G G
39896 gttgtgcttttattcagttaaagagatttttaggtattttacagat
16 V V L Y S V K E I F R Y F T D
39941 tctaacttacaacgtaaaaaatcaatttagaacaatatatccg
31 S N L Q R K K I N L E Q I Y P
39986 atatatttagattgttttaaaaggctaaaaagatgattggagct
15 46 I Y L D C F K K A K K M I G A
40031 tatattattccaacagacagcatgaatttttagattttttgat
61 Y I I P T E Q H E F L D F F D
40076 attgaagtctttaataatttagataagcaaagtaaaaaagcgtat
76 I E V F N N L D K Q S K K A Y
40121 gaaaatggtattggatttagacaaatgattaatttatcaaataga
20 91 E N V I G F R Q M I N L S N R
40166 gttaaggcaatggaagattttaagatgagtttcaacaatgaattt
106 V K A M E D F K M S F N N E F
40211 agtacaaatcagatttttttaaatccttcttttggtatggaaca
121 S T N Q I F F N P S F V M E T
25 40256 attgctattataaatgaatatcaaaaagatatatcttatttaaaa
136 I A I I N E Y Q K D I S Y L K
40301 aatataattaataaaaatgaatgaaaatagagcttataatcatatt
151 N I I N K M N E N R A Y N H I
40346 gatagttttatcacttcagagtaccgacgaaaaataaacgattat
166 D S F I T S E Y R R K I N D Y
30 40391 aatctttatcttgataaatttgaagaacagtttagtcaaaagttt
181 N L Y L D K F E E Q F S Q K F
40436 aaaataaacagaacttcgataaaaagaagaattattattaattta
196 K I N R T S I K E R I I I N L
40481 aacaagaggagatttaaatga 40501
35 211 N K R R F K *

40

45

50

55

Physico-chemical parameters of ORF 77ORF019

1 MNEQIIGSIY TLAGGVVLYS VKEIFRYFTD SNLQKKINL EQIYPIYLDG FKKAKKMGQA
 61 YIIPTQEHF LDFPDIEVPN NLDKQSKAY ENVIGFRQMI NLSNRVKAME DFMSPNNEP
 121 STNQIFFNPS FVMETIAIIN EYQDISYLYK NIINKGNENR AYNHIDSFIT SEYRRKINDY
 181 NLYLDKFEEQ FSQKFKINRT SIKERIIINL NKRRFK

Number of amino acids: 216
 Average molecular weight (Daltons): 26026.06
 Mean amino acid weight (Daltons): 120.49
 Monoisotopic molecular weight (Daltons): 26009.34
 Mean amino acid monoisotopic weight (Daltons): 120.41

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	7	3.24%	7.58%	Cys	C	1	0.46%	1.66%
Asp	D	10	4.63%	5.28%	Glu	E	16	7.41%	6.37%
Phe	F	19	8.80%	4.09%	Gly	G	5	2.31%	6.84%
His	H	2	0.93%	2.24%	Ile	I	28	12.96%	5.81%
Lys	K	22	10.19%	5.95%	Leu	L	12	5.56%	9.42%
Met	M	7	3.24%	2.37%	Asn	N	23	10.65%	4.45%
Pro	P	3	1.39%	4.9%	Gln	Q	10	4.63%	3.97%
Arg	R	11	5.09%	5.16%	Ser	S	13	6.02%	7.12%
Thr	T	7	3.24%	5.67%	Val	V	7	3.24%	6.58%
Trp	W	0	0.00%	1.23%	Tyr	Y	13	6.02%	3.18%

Number of acidic (negative) amino acids (ED): 26
 12.04%
 Number of basic (positive) amino acids (KR): 33
 15.28%
 Total charge (KRED): 59
 27.31%
 Net charge (KR - ED): 7
 3.24%
 Theoretical pI: 9.52
 Total linear charge density: 0.28
 Average hydrophobicity: -4.84
 Ratio of hydrophilicity to hydrophobicity: 1.37
 Percentage of hydrophilic amino acid: 54.17%
 Percentage of hydrophobic amino acid: 45.83%
 Ratio of %hydrophilic to %hydrophobic: 1.18

5

157

77ORF043 sequence

10

15

20

25

30

35

40

45

50

55

```
29304   atgtattacgaaataggcgaaatcatacgcaaaaatattcatgtt
1    M Y Y E I G E I I R K N I H V
29349   aacggattcgattttaagctattcatttttaaagggtcatatgggc
16   N G F D F K L F I L K G H M G
29394   atatcaatacaagttaaagatatgaacaacgtaccaattaaacat
31   I S I Q V K D M N N V P I K H
29439   gcttatgtcgtagatgagaatgacttagatatggcatcagactta
46   A Y V V D E N D L D M A S D L
29484   ttttaaccaagcaatagatgaatggattgaagagaacacagacgaa
61   F N Q A I D E W I E E N T D E
29529   caggacagactaattaacttagtcatgaaatggtag 29564
76   Q D R L I N L V M K W *
```

Physico-chemical parameters of ORF 77ORF043

1 MYRIGEIR KNHVNFGDFP KLFILKGHMG ISIQVKDMNN VPIKHAYVVD ENLDMASDL
61 FNQAIDEWIE ENTDEQDRLL NLVMKW

Number of amino acids: 86
Average molecular weight (Daltons): 10186.68
Mean amino acid weight (Daltons): 118.45
Monoisotopic molecular weight (Daltons): 10180.02
Mean amino acid monoisotopic weight (Daltons): 118.37

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	3	3.49%	7.58%	Cys	C	0	0.00%	1.66%
Asp	D	9	10.47 %	5.28%	Glu	E	7	8.14%	6.37%
Phe	F	4	4.65%	4.09%	Gly	G	4	4.65%	6.84%
His	H	3	3.49%	2.24%	Ile	I	11	12.79 %	5.81%
Lys	K	6	6.98%	5.95%	Leu	L	6	6.98%	9.42%
Met	M	5	5.81%	2.37%	Asn	N	8	9.30%	4.45%
Pro	P	1	1.16%	4.9%	Gln	Q	3	3.49%	3.97%
Arg	R	2	2.33%	5.16%	Ser	S	2	2.33%	7.12%
Thr	T	1	1.16%	5.67%	Val	V	6	6.98%	6.58%
Trp	W	2	2.33%	1.23%	Tyr	Y	3	3.49%	3.18%

Number of acidic (negative) amino acids (ED): 16
18.60%
Number of basic (positive) amino acids (KR): 8
9.30%
Total charge (KRED): 24
27.91%
Net charge (KR - ED): -8
9.30%
Theoretical pI: 4.38
Total linear charge density: 0.30
Average hydrophobicity: -2.80
Ratio of hydrophilicity to hydrophobicity: 1.19
Percentage of hydrophilic amino acid: 48.84%
Percentage of hydrophobic amino acid: 51.16%
Ratio of %hydrophilic to %hydrophobic: 0.95

5

159

77ORF102 sequence

29051 atgagcaacatttataaaaagctacctagtagcagtattatgcttc
1 M S N I Y K S Y L V A V L C F
10 29096 acagtcttagcgattgtacttatgccgtttctatacttcactaca
16 T V L A I V L M P F L Y F T T
29141 gcatgggtcaattgcgggattcgcaagtatcgcaacattcatgtac
31 A W S I A G F A S I A T F M Y
15 29186 tacaagaatgctttttcaaagaataa 29212
46 Y K E C F F K E *

20

25

30

35

40

45

50

55

Physico-chemical parameters of ORF 77ORF102

1 MSNIYKSYLV AVLCPVLAI VIMPFLYFTT AWSIAGFASI ATPMYKCEP FKE

Number of amino acids: 53
 Average molecular weight (Daltons): 6155.42
 Mean amino acid weight (Daltons): 116.14
 Monoisotopic molecular weight (Daltons): 6151.07
 Mean amino acid monoisotopic weight (Daltons): 116.06

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	6	11.32 %	7.58%	Cys	C	2	3.77 %	1.66%
Asp	D	0	0.00%	5.28%	Glu	E	2	3.77 %	6.37%
Phe	F	7	13.21 %	4.09%	Gly	G	1	1.89 %	6.84%
His	H	0	0.00%	2.24%	Ile	I	4	7.55 %	5.81%
Lys	K	3	5.66%	5.95%	Leu	L	5	9.43 %	9.42%
Met	M	3	5.66%	2.37%	Asn	N	1	1.89 %	4.45%
Pro	P	1	1.89%	4.9%	Gln	Q	0	0.00 %	3.97%
Arg	R	0	0.00%	5.16%	Ser	S	4	7.55 %	7.12%
Thr	T	4	7.55%	5.67%	Val	V	4	7.55 %	6.58%
Trp	W	1	1.89%	1.23%	Tyr	Y	5	9.43 %	3.18%

Number of acidic (negative) amino acids (ED): 2
 3.77%
 Number of basic (positive) amino acids (KR): 3
 5.66%
 Total charge (KRED): 5
 9.43%
 Net charge (KR - ED): 1
 1.89%
 Theoretical pI: 8.18
 Total linear charge density: 0.13
 Average hydrophobicity: 10.81
 Ratio of hydrophilicity to hydrophobicity: 0.40
 Percentage of hydrophilic amino acid: 28.30%
 Percentage of hydrophobic amino acid: 71.70%

WO 00/32825

PCT/IB99/02040

5

161

Ratio of %hydrophilic to %hydrophobic:

0.39

10

15

20

25

30

35

40

45

50

55

5

162

77ORF104 sequence

10

15

20

25

30

35

40

45

50

55

```
34393      atggtaaccaagaatttttaaaaaactaaacttgagtgttcagat
1   M   V   T   K   E   F   L   K   T   K   L   E   C   S   D
34438      atgtacgctcagaaactcatagatgaggcacagggcgatgaaaat
16  M   Y   A   Q   K   L   I   D   E   A   Q   G   D   E   N
34483      aggttgtagacctaatttatccaaaaacttcagaaacggtcataca
31  R   L   Y   D   L   F   I   Q   K   L   A   E   R   H   T
34528      cgccccgctatcgtcgaatattaa 34551
46  R   P   A   I   V   E   Y   *
```

Physico-chemical parameters of ORF 77ORF104

1 MVTKEFLKTK LECSDMYAQK LIDEAQGDEN RLYDLPIQKL AERHTRPAIV EY

Number of amino acids: 52
 Average molecular weight (Daltons): 6193.13
 Mean amino acid weight (Daltons): 119.10
 Monoisotopic molecular weight (Daltons): 6189.12
 Mean amino acid monoisotopic weight (Daltons): 119.02

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	4	7.69 %	7.58%	Cys	C	1	1.92%	1.66%
Asp	D	4	7.69 %	5.28%	Glu	E	6	11.54 %	6.37%
Phe	F	2	3.85 %	4.09%	Gly	G	1	1.92%	6.84%
His	H	1	1.92 %	2.24%	Ile	I	3	5.77%	5.81%
Lys	K	5	9.62 %	5.95%	Leu	L	6	11.54 %	9.42%
Met	M	2	3.85 %	2.37%	Asn	N	1	1.92%	4.45%
Pro	P	1	1.92 %	4.9%	Gln	Q	3	5.77%	3.97%
Arg	R	3	5.77 %	5.16%	Ser	S	1	1.92%	7.12%
Thr	T	3	5.77 %	5.67%	Val	V	2	3.85%	6.58%
Trp	W	0	0.00 %	1.23%	Tyr	Y	3	5.77%	3.18%

Number of acidic (negative) amino acids (ED): 10
 19.23%
 Number of basic (positive) amino acids (KR): 8
 15.38%
 Total charge (KRED): 18
 34.62%
 Net charge (KR - ED): -2
 3.85%
 Theoretical pI: 5.03
 Total linear charge density: 0.38
 Average hydrophobicity: -5.81
 Ratio of hydrophilicity to hydrophobicity: 1.47
 Percentage of hydrophilic amino acid: 53.85%
 Percentage of hydrophobic amino acid: 46.15%

WO 00/32825

PCT/IB99/02040

5

164

Ratio of %hydrophilic to %hydrophobic:

1.17

10

15

20

25

30

35

40

45

50

55

5

77ORF182 sequence

10

15

20

25

30

35

40

45

50

55

```
29268      atgttcaatataaaacgaaaaacggaggaagtcaagatgtattac
1   M F N I K R K T E E V K M Y Y
29313      gaaataggcgaaatcatacgcaaaaatattcatgttaacggattc
16  E I G E I I R K N I H V N G F
29358      gatTTtaagctattcatttttaaaggTcatatgggcatatcaata
31  D F K L F I L K G H M G I S I
29403      caagttaaagatatgaacaacgtaccaattaacatgcttatgtc
46  Q V K D M N N V P I K H A Y V
29448      gtagatgagaatgacttagatatggcatcagacttatttaaccaa
61  V D E N D L D M A S D L F N Q
29493      gcaatagatgaatggattgaagagaacacagacgaacaggacaga
76  A I D E W I E E N T D E Q D R
29538      ctaattaacttagtcatgaaatggtag 29564
91  L I N L V M K W *
```

Physico-chemical parameters of ORF 77ORF182

1 MFNIKRKTEE VMYYEIGEI IRKNIHVNGF DFKLFILKGH MGISIQVKDM NNVPIKHAYV
 61 VDENDLDMAS DLFNQAIDEW IEENTDEQDR LINLVMKW

Number of amino acids: 98
 Average molecular weight (Daltons): 11691.50
 Mean amino acid weight (Daltons): 119.30
 Monoisotopic molecular weight (Daltons): 11683.84
 Mean amino acid monoisotopic weight (Daltons): 119.22

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	3	3.06 %	7.58%	Cys	C	0	0.00%	1.66%
Asp	D	9	9.18 %	5.28%	Glu	E	9	9.18%	6.37%
Phe	F	5	5.10 %	4.09%	Gly	G	4	4.08%	6.84%
His	H	3	3.06 %	2.24%	Ile	I	12	12.24 %	5.81%
Lys	K	9	9.18 %	5.95%	Leu	L	6	6.12%	9.42%
Met	M	6	6.12 %	2.37%	Asn	N	9	9.18%	4.45%
Pro	P	1	1.02 %	4.9%	Gln	Q	3	3.06%	3.97%
Arg	R	3	3.06 %	5.16%	Ser	S	2	2.04%	7.12%
Thr	T	2	2.04 %	5.67%	Val	V	7	7.14%	6.58%
Trp	W	2	2.04 %	1.23%	Tyr	Y	3	3.06%	3.18%

Number of acidic (negative) amino acids (ED): 18
 18.37%
 Number of basic (positive) amino acids (KR): 12
 12.24%
 Total charge (KRED): 30
 30.61%
 Net charge (KR - ED): -6
 6.12%
 Theoretical pI: 4.76
 Total linear charge density: 0.33
 Average hydrophobicity: -3.89
 Ratio of hydrophilicity to hydrophobicity: 1.28

5

167

Percentage of hydrophilic amino acid:
Percentage of hydrophobic amino acid:
Ratio of %hydrophilic to %hydrophobic:

51.02%
48.98%
1.04

10

15

20

25

30

35

40

45

50

55

Table 5

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100017|lan|77ORF017 Phage 77 ORF |23269-23982|-3
(237 letters)

Database: nr

393,678 sequences; 120,452,765 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 4493986 emb CAB39045.1 (AL034559) predicted using hexExon; ...	41	0.010
gi 730607 sp P23250 RPI1_YEAST NEGATIVE RAS PROTEIN REGULATOR P...	38	0.053
gi 3097044 emb CAA75299 (Y15035) K1R (Cowpox virus)	38	0.090
gi 2146245 pir S73794 hypothetical protein H91_orf180 - Mycopl...	38	0.090
gi 83910 pir S04682 ribosomal protein var1 - yeast (Candida gl...	37	0.15
gi 133135 sp P21358 RMAR_CANGA MITOCHONDRIAL RIBOSOMAL PROTEIN ...	37	0.15
gi 2128843 pir H64475 hypothetical protein MJ1409 - Methanococ...	36	0.20
gi 5107017 gb AAD39926.1 AF126285_2 (AF126285) RNA polymerase (...)	36	0.35
gi 2146210 pir S73342 hypothetical protein E07_orf166 - Mycopl...	35	0.60

Database: swissprot

79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P23250 RPI1_YEAST NEGATIVE RAS PROTEIN REGULATOR PROTEIN.	38	0.014
sp P21358 RMAR_CANGA MITOCHONDRIAL RIBOSOMAL PROTEIN VAR1.	37	0.040
sp Q21444 LDLC_CAEL LDLC PROTEIN HOMOLOG.	34	0.35
sp P27240 RFAY_ECOLI LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS PROT.	33	0.46
sp P53192 YGCO_YEAST HYPOTHETICAL 27.1 KD PROTEIN IN ALK1-CKB1.	33	0.60
sp P32908 SMC1_YEAST CHROMOSOME SEGREGATION PROTEIN SMC1 (DA-B.	33	0.60
sp P54683 TAGB_DICDI PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR .	32	0.78
sp Q03100 CYAA_DICDI ADENYLATE CYCLASE, AGGREGATION SPECIFIC (.	32	0.78

169

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100019|lan|77ORF019 Phage 77 ORF|39851-40501|2
(216 letters)Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341966 dbj BAA31932 (AB009866) orf 59 (bacteriophage phi PVL)	437	e-122
gi 2689911 (AB000792) B. burgdorferi predicted coding region BB...	38	0.058
gi 1171589 emb CAA64574 (X95275) frameshift (Plasmodium falcip...	37	0.10
gi 4493986 emb CAB39045.1 (AL034559) predicted using hexExon; ...	36	0.23
gi 141257 sp P18019 YPI9_CLOPE HYPOTHETICAL 14.5 KD PROTEIN (OR...	36	0.29
gi 133412 sp P27059 RPOB_ASTLO DNA-DIRECTED RNA POLYMERASE BETA...	35	0.51
gi 3122231 sp Q58851 HISX_METJA HISTIDINOL DEHYDROGENASE (HDH) ...	35	0.51
gi 3649757 emb CAB11106.1 (Z98547) predicted using hexExon; MA...	34	0.66
gi 2688313 (AB001146) sensory transduction histidine kinase, pu...	34	0.87

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P18019 YPI9_CLOPE HYPOTHETICAL 14.5 KD PROTEIN (ORF9).	36	0.079
sp Q58851 HISX_METJA HISTIDINOL DEHYDROGENASE (EC 1.1.1.23) (H.	35	0.14
sp P27059 RPOB_ASTLO DNA-DIRECTED RNA POLYMERASE BETA CHAIN (E.	35	0.14
sp Q02224 CENE_HUMAN CENTROMERIC PROTEIN E (CENP-E PROTEIN).	34	0.31
sp P04931 ARP_PLAFA ASPARAGINE-RICH PROTEIN (AG319) (ARP) (FRA...	33	0.53
sp P18011 IPAB_SHIFL 62 KD MEMBRANE ANTIGEN.	32	0.69
sp P18709 VTA2_XENLA VITELLOGENIN A2 PRECURSOR (VTG A2) [CONTA...	32	0.90
sp Q64409 CP3H_CAVPO CYTOCHROME P450 3A17 (EC 1.14.14.1) (CYPI...	32	0.90
sp P21358 RMAR_CANGA MITOCHONDRIAL RIBOSOMAL PROTEIN VARI.	32	0.90
sp Q03945 IPAB_SHIDY 62 KD MEMBRANE ANTIGEN.	32	1.2

170

BLASTP 2.0.8 (Jan-05-1999)

Query= sid|100043|lan|77ORF043 Phage 77 ORF|29304-29564|3
(86 letters)Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341947 dbj BAA31913 (AB009866) orf 39 (bacteriophage phi PVL)	182	6e-46
gi 744518 prf 2014422A FKBP-rapamycin-associated protein (Homo...	32	0.84
gi 1169736 sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN...	32	0.84
gi 1169735 sp P42345 FRAP_HUMAN FKBP-RAPAMYCIN ASSOCIATED PROTE...	32	0.84
gi 3282239 (U88966) rapamycin associated protein FRAP2 (Homo sa...	32	0.84
gi 3875402 emb CAA98122 (Z73906) cDNA EST EMBL:D64544 comes fr...	31	2.5
gi 1084792 pir S54091 hypothetical protein YPR070w - yeast (Sa...	30	4.2

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P42345 FRAP_HUMAN FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP)	32	0.24
sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP) (R.	32	0.24
sp P34554 YNP1_CAEEL HYPOTHETICAL 42.2 KD PROTEIN T05G5.1 IN C.	28	3.5
sp Q24118 LIO DROME LINOTTE PROTEIN.	28	3.5
sp P80034 ACH2_BOMMO ANTICHYMOTRYPSIN II (ACHY-II).	28	3.5
sp P22922 ALAT_BOMMO ANTITRYPSIN PRECURSOR (AT).	28	3.5
sp Q44363 TRAA_AGR76 CONJUGAL TRANSFER PROTEIN TRAA.	28	3.5
sp P38255 YBU5_YEAST HYPOTHETICAL 51.3 KD PROTEIN IN PHO5-VPS1.	27	6.0
sp P55822 SH3B_HUMAN SH3BGR PROTEIN (21-GLUTAMIC ACID-RICH PRO.	27	7.9
sp Q58482 YAG2_METJA HYPOTHETICAL PROTEIN MJ1082.	27	7.9
sp P34252 YKK8_YEAST HYPOTHETICAL 52.3 KD PROTEIN IN HAP4-AAT1.	27	7.9

5

171

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100102|lan|77ORF102 Phage 77 ORF|29051-29212|2
(53 letters)

10

Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341946 dbj BAA31912 (AB009866) orf 38 [bacteriophage phi FVL]	96	3e-20
gi 4325288 gb AAD17315 (AF123593) voltage-dependent sodium cha...	28	7.1
gi 2649684 (AE001040) A. fulgidus predicted coding region AF092...	28	9.3

15

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P42087 HUTM_BACSU PUTATIVE HISTIDINE PERMEASE.	26	7.1
sp P04775 CIN2_RAT SODIUM CHANNEL PROTEIN, BRAIN II ALPHA SUBU...	26	9.2
sp P42619 YQJF_ECOLI HYPOTHETICAL 17.2 KD PROTEIN IN EXUR-TDCC...	26	9.2

20

25

30

35

40

45

50

55

172

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100104|lan|77ORF104 Phage 77 ORF|34393-34551|1
(52 letters)Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 2315523 (AF016452) similar to the leucine-rich domains found...	29	4.2
gi 4377168 gb AAD18990 (AE001666) CT711 hypothetical protein [...]	29	5.4
gi 3882171 dbj BAA34445 (AB018268) KIAA0725 protein [Homo sapi...]	28	9.3

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P04879 RRPP_VSVIG RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48.	27	5.4
sp P04880 RRPP_VSVIM RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48.	27	5.4
sp Q13946 CN7A_HUMAN HIGH-AFFINITY CAMP-SPECIFIC 3',5'-CYCLIC	26	7.1
sp P35381 ATPA_DROME ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL P.	26	9.3
sp P54659 MVPB_DICDI MAJOR VAULT PROTEIN BETA (MVP-BETA).	26	9.3
sp P40397 YHXC_BACSU HYPOTHETICAL OXIDOREDUCTASE IN APRE-COMK	26	9.3

5

173

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|122748|lan|77ORF182 Phage 77 CRF|29268-29564|3
(98 letters)

10

Database: nr
393,678 sequences; 120,452,765 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341947 dbj BAA31913.1 (AB009866) orf 39 (bacteriophage phi..	182	8e-46
gi 1084792 pir S54091 hypothetical protein YPR070w - yeast (Sa..	35	0.13
gi 1169736 sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN..	32	1.1
gi 744518 prf 2014422A FKBP-rapamycin-associated protein (Homo..	32	1.1
gi 5051381 emb CAB44736.1 (AL049653) dJ647M16.2 (FK506 binding..	32	1.1
gi 4826730 ref NP_004949.1 pFRAP1 FK506 binding protein 12-rap..	32	1.1
gi 3282239 (U88966) rapamycin associated protein FRAP2 (Homo sa..	32	1.1

15

Database: swissprot
79,909 sequences; 29,054,478 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P42345 FRAP_HUMAN FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP) .	32	0.29
sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP) (R.	32	0.29
sp P40557 YIA5_YEAST PUTATIVE DISULFIDE ISOMERASE YIL005W PREC.	29	3.3
sp Q24118 LIO_DROME LINOTTE PROTEIN.	28	4.4
sp Q44363 TRAA_AGRT6 CONJUGAL TRANSFER PROTEIN TRAA.	28	4.4
sp P80034 ACH2_BOMMO ANTICHYMOTRYPSIN II (ACHY-II).	28	4.4
sp P34554 YNP1_CAEEL HYPOTHETICAL 42.2 KD PROTEIN T05G5.1 IN C.	28	4.4
sp P22922 ALAT_BOMMO ANTITRYPSIN PRECURSOR (AT).	28	4.4

25

30

35

40

45

50

55

Table 6

1st position (5' end)	2nd position				3rd position (3' end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

5

175

Table 7

10

Bacteriophage 3A, complete genome sequence

15

20

25

30

35

40

45

50

55

```

1  caaacgctag caacgcggat aaatttttca tgaaggggg tcttttatg aagttacaa aaaaacagct
71  aaaaagatgt atagaaagatt acaaaaaaac cgaatgacata ttaattaatt tgtatataga aacatgatga
141  tttttattgtc gggttaagaga tgaactttaa aatagtattt taatgataga gcatacaaac aaggctgggtg
211  cgcgcgaatgt tattaagaaat ccatttaagca tagaactgac aaaaacagct caaacactaa ataacttact
281  caagtcttatg gggttaactgt cagcacaagag aaaaagataa gttcaagaag aaggtggatt cggtgactat
351  taaagtttta aatgaacctt caccaaaaact attacaaca tggatgcag agcaagtca ctaaggggaaa
421  ataaaaacaa gcaaatatgt tagaaaagaa tggagagac atcttagata tctagaaat ggaggttaast
491  ggggtattga tgaagaatta gcgcacgcgc ctattcgatt tatagaaaag tttgttaaac ctccaaagg
561  attcaaacgt caactgtgat tacagccatg gcaacatttt attatcgcca gtttgggttg ttgggtttcat
631  aagaagacaa aactgcgcag gtttaagaa gctttgatgt ttatggggcg aaaaatgggt aaaaacacca
701  ctattctctg ggttgcgaac tatgtgtgat cacaagatgg agaaatgggt gcagaanaatc atttgttagc
771  aaacgttaag aacacagcta ggattctatt tgaatgaact aaggcgatga ttaagagtag cccaaagctt
841  gataaaaatt tcagaaacatt aagagatgaa atccattatg acgcacagat atcaaaaatt atgccccag
911  catcagatag cgataagtta gatggattga atcacacat ggggtttttt gatgaatttc atgaatttaa
981  agcatataaa ttgatttcag ttataaaaaa ctcaagagct gcaaggttac aacctcttct catctacatt
1051  agcgcgcagc ggtatcaatt agatgggtcca ctgttgata tggtagaagc gggagagagc accttagatc
1121  aactcataga agagaaagaa actttttatt atttagcacc tttggatgat gaogatgata ttaattgattc
1191  gtcgaactgt ataaaagcaa atccccactt aggtgtctct ataaatttag atgagatgaa agaaagattg
1261  gaaaaagcta agagaacacc agctgaacgt ggagatttta taaccaaaag gtttaataatc ttctgaata
1331  atgacagatg gatttttatt gattacccaa cactccaaaa aaatgaatgaa attgtttctt tagaagagct
1401  ggaaggcaga ccgtgcacga ttggttatga tttatcagaa acagaggact ttacagccgc gtgtgctact
1471  tctgcgttag ataaggttaa agttgcagtt ttatcgcat catggattcc taagcacaana gttgaatatt
1541  tcaacgaaaa aataccctat agagaatggg aagaagatgg ctatttaaca gtgcaagata agccttatat
1611  tgactaccan gatgttttaa attggataat taagatgant gacattatgt tagtagaana aattacttat
1681  gatagagoga acgcattcaa actnaatcaa gatttaaaaa attacgggtt tgaacgggaa gaacaaagc
1751  aaggagcttt gaccttgagc cctgcattga aggtatttaa agaaatgttt ttagatggga aaataatatt
1821  taataataat cctttaatga aatggatat cantaatgtt cagttgaaac tagacagaaa cggaaactgg
1891  ttgcgctcta agcaaacgag atatcgtaaa atagatggct ttgcagcatt ttaaacaca tatcacagta
1961  ttatgaataa agttgtttct gatagtggtg aaggaaacat agagtttatt agtattaaag acataatgcg
2031  ttaaggaggt gaattgtatc gcaaaaagaa atattgtcac acgcataaag aanaaatga tagacaattg
2101  ttatgatcag tcaacttcta agctttatga ctttagccca tggaaaaata gatctttttg ggggttaatt
2171  aataatacgc ttgaanaata tgaacagata ttttcagcta ttcaaaagt atcttaattc atggcttagtt
2241  tgcctttgaa aatgatgaa gattataaag tagttaatca agaatgtctt gatttactta cagtgctacc
2311  gaataactct ctgagcagct ttgattttat taatcaaaat gaacaaatca gaattgaaaa aggttaatgca
2381  tatgtgctaa ttgaacgaga catctatcat caacctcaaa agcttttctt attaatcca gatgttggtt
2451  aaattgttaa tgaanaacaa tcaactgaaac ttattatttc catctatgct gcaactggaa ataatgtat
2521  ttatgcataa atggacatgt tgcattttaa acacatcgtg gcacttaata tggtagcaag cattagtccg
2591  attgatgtgt tgaagaatac aactgatttt gataatgcag taagaacctt taactttaca gaattgaaa
2661  aactgtattc ttctatgctt aaatatgggt ccaatgtagg taagaanaaa aggcagcaag gtttagaaga
2731  ttcaaacag tactatgaag aaaaacgggt aatattattc caagagcctg gtgttgaaat cgaacccgta
2801  cctaaaaaat atgtctctga agatatagtg gcaagcgaga atttaacaag agaaagugto gctaacgctt
2871  ttcaattgcc ctacgatttc ttaaatgcaa gatcaaatc aaatttcgcg aaaaatgaag agttaaacag
2941  attttacttg cagcatacct tattgccaat cgtcaaacag tatgaagaag aatttaacg gaaactacct
3011  actaaacag acagagaaaa aaatgggtat tttaattta acgttaaatc ttatttaagg gctgatagtg
3081  caacacaaagc agaaggttac ttaaaagcag ttctgtatgg ttactacact ataagtaca tttagaggtg
3151  ggaagattta ccaccagttg aaggtggaga taagccgcta atagccggt atttataccc aattgacagc
3221  ccacttgaaat taagaaaaatc ttgaaaggt ggtgataaaa atgtcaatga agtcaagta ttttcaaatg
3291  aaaaagaaat caaaaagtaa aggtgaataa ttattttatg gtgatattgt aagtgaataa tgggttgaaa
3361  gtgatgtaac tgcacagat ttcaaaaata aactagatga actaggagac atcagtgaan tagatgttca
3431  tataaatcca tctggagcca gtgtatttga agggatgcca atatacaata tgcataaaat gcactcgtca
3501  aaaaataata tctatgtcga tgccttagcg gcatacaatt ctagtgttat cgtatgagt ggtgacacta
3571  tttttatgca caaaaatagt tttttaatga ttcatattc atgggttatg actgtaggt atgcagaaga
3641  gtttaagaaag acagcggatt tacttgaana acagatgct gttagtattc cagcttattt agataaagc
3711  aaagatttag atcaagaaac tttaaacag atgttagatg cagaactgtt gcttactgca gaagaagctt
3781  tgtctcttgg ctgtacagat gaaattttag gagttaatga ataaactgct agtatctcta angagcaata
3851  taagcgtttc ggaacgttcc cagaagattt aagaaagat gtagacaaa tcaactaaaat cgtatgata
3921  gatagtttg aattgggtga aaacccatga aaagatgtg cactagaga aaaaagaaaa agagaaaaaa
3991  ttaaacgaga atgcgaatt ttanaaatga caatgaatg ttaggaggaa atgaatgac gacattatat
4061  gaattaaaaa aatccttagg tatgattgga caacatttaa aaataaaaa tgaatgaatt agtcagaag
4131  caacagaccc aattattgat atggagaca tcaaaaact agaaacaga aagcagggt tacaacaaag
4201  atttaacatt gttgaagac aagtaaaaga cattgaaga aagaaaaaag cgaagttaa agacacagga
4271  gaagctttac aatctttaa cgalcatgag aagatggta aagctaaagg agagtttat cgtcacgga
4341  ttttccaaa tgaattgaa aaaccttcaa tggagycaca acgtttatta cagcgtttac caacaggtaa
4411  tgattcagg ggtgaagaag ccttaccaaa aacactttct aaagaaattg tttcagaacc atttgcataa
4481  aaccaattac gtgaaaaagc cgttcaact aacattaaag gtttagagat tccaaagatt tcatatactt
4551  tgaacgatga tgaattcatt acagatgcag aaacagcaaa agaatnaaa ttaaaaggtg atacagttaa
4621  attcactact aataaattca aagatttgc tgaatttca gatactgaa ttcatggatc agatgatgat
4691  ttagtaaaact ggttgaaaa cgcactacaa tcaggtctag cagctaaaga acgtaaagat gccttagcag

```

5

176

10

15

20

25

30

35

40

45

50

55

4761 taagtccata atctggatta gatcacatgt ctttttcaaa tggatctgtt aaagaagttg agggagcaga
4831 cctgtatgat gctattatta agcttttggc agattttacat gaagattacc gtgataacgc aacaatttat
4901 atgcgatatg cggattatgt caaaattcatt agtgttcttt caantggaa acaaatattc tttgacacac
4971 cagcagaaaa agcatttggc aaaccagtag tttttacaga tgcagcagtt aaacctattg tggagatttt
5041 caattatttt ggaatttaact atgatggaa acattatgac actgataaag atgttaaaaa agggagaaat
5111 ttgtttgtat taactgcatg gtatgatcag caacgtacat tagacagttc attcagaatt gcaaaagcaa
5181 aagaaaaaac aggttcatta cccagctaa ccccaaaagg ttaattgaac agtcaaggtt aaatcagctg
5251 taatatcagc cgaatagggg tgatgaatg agtttagaag aaattaaatt gtggttgaga attgactata
5321 atttcgaaaa tgatttaatt gaaggtctca ttcactcggc taagtctgaa ttactattaa gtggggttcc
5391 agattatgac aaagatgact tggaaatccc gcttttttgt acagcgatta gatatatcat tgcagagat
5461 tatgaagtc gtgggtactt aaatgaccaa tctagaagca aggtttttaa tgaagggga ttgcataaaa
5531 tgattctgaa attaaaaaag tggtaggtga tttttaaatg gaatttaatt aatttaaga tgcgcatat
5601 ttttttcaat atgtanaata agggcgttat ccagatgaag aggaaaaaat gaagtgtat agttgctttt
5671 gcaaatata caatcttct atgaagata gagaattttt aaagcgact gaatcaagt caggataaac
5741 cataattatg aggtcttcta aaattgaata tctaccacaa acaaatcact tagttaaat tgcagagggc
5811 ttatattccg ataaattatt caacattaaa gaataaagaa ttgatacacc agtatattgc tataatcacg
5881 tggttttatc agaaaaatga gtgtagaat taaggggata cctgaaggtt tgaagaattt agaatcggta
5951 tgcgttaaac aatcaatgca agtaagagt gatagagctt taatgaagc atctgaattt tcttaagg
6021 ctttaaaaga agaatctgag agttttaag atacgggtgc tagcatagaa gaataacta aatcaagcc
6091 ttatcacaaa taggaagtc aagaaagagc tgttttaatt gaattgggtg gccctatgaa tgcataaaac
6161 attattcact tgatgaaca tgggtatata agatagga aaataatata accaagagt ttrggagtta
6231 ttgcaaaaac attagctgct aatgaacgga agtatagaga aattataaaa aaggagtgg ccagataaat
6301 gaataatata aacacataaa aagaattttt attatctgat gcagagctcc aaactatata aactctaga
6371 attactactt ataaagtac tgaataatgt gaatttcca aacttttgt tgtattaca cctattatg
6441 attactcttc agactctatg tctgataaat atcttagtga agaatctta attcaaatag atgtagaatc
6511 ttcaaatat cagaaaaaaa ttgataaac aaacgaaata agatattgt tatatacaaa aaattcaatt
6581 caagcatcta gttagttaga tggatttttt gaagaaacta aacgtctat gatgtcaga cgttatcaag
6651 ttatcacaaa aatatatat tataaaaaac agcgatcga ataggtctgc ttttaattt ttaaggagga
6721 aataagcaat ggcagaagga caaggttctt ataaagttag ttttaaaaga ttatagttg pagtttttaa
6791 cccagagaga acaaaagtag ttaaacgcat gacatggaaa gatgaanaag gtggtacagt tgattcaaat
6861 atccaggtt tagcaccaga tttagtagat atgtttgcat ctacaaaacg tgtttgagat aaaaaaacag
6931 gtactaata agttaagctt gacatgata ttttaaatat tccaagtga gatctaaata cagtatttgg
7001 tctgtctaaa gataaaaaat gtacattctg ggtaggagag aatacaagag caccatagct aacagttatt
7071 ggaagatctg aagatggttt aacaggtcaa ccagtgtag ttcgctact taagggtact tttagcttgg
7141 tcttcaatga atttaaaaca cggaggagaa aagcagaagc accagagcca acaaaaattaa ctggtgactg
7211 gtagacaga aagtgatgt tgatggatc tccacaaggt attgtatag ggtatcatga agttaagaa
7281 ggaagagcag aattcttcaa aaagtattc gttgataca cggacagta agatcattca aggtattctg
7351 caagtctgct acccagctaa ccccaaaatt gttgaagtag cagttaattc aaatctgca acagtttccg
7421 cagaataggg gctttcaaaa taactcaag gagaataatt tatgactaaa actttaagg tttataaagg
7491 agacagagtc gtatgtctg acaaggtga aggcanaagtc tcaatattt tatctaat tgaagcggat
7561 caactttac caaaggtac ttaccaagtc gctgtggag aaatggtta agaatctat aagttgatg
7631 taactcaatt caaaaccaat ccaattctag tctcagggtt atcatttaca cccgaacta aatcaatcac
7701 ggttaaatgt gatgaacat tgaaacaaa cactgcacca agtacagcaa cgaataaaac gttgaatat
7771 accagtgaa atccagagtt tgttactggt gatgagagaa caggagcaat tcaaggtga gctgaggaa
7841 tctcagttat cactgctacg tctactgacg gaagtgcaca gctcggacaa attacagtaa cagtaacaaa
7911 tggataatta tttagagcgc agaataatct cgtctttttt atttgaataa aaggagctaa tacaatgatt
7981 aaatttgaaa ttaagagcgc taaaacgga aaaaacagga gctatataaa aagaagatgt acaatgggag
8051 aagcagaaa atgctatgag tatttagaat tagtaaatca agagaataaa aagaagatc ctacagcaac
8121 aaaaatgaga caaaaagagc gacagttatt agtagattta tttaaagatg aaggattgac tgaagagat
8191 gttttgaaca agatgagcac taactttat acaaaagctt tgaagatat atttcagaa atcaatgggtg
8261 aagatgaaga agattcagaa actgaaccag aagagatggg aaagacagaa gaacaaatct aataaagat
8331 attttatcga acattaaaga aatcacacgt tctgtatagg agcagtatgg gtggacatta actgaagtc
8401 gaaaaacagcc gtatgcataa ctttttagaaa tacttaatga agagaataaa gaagagactg aagaaaaaca
8471 aagtgaacaa aagtcattca caggtacgga ctttaaaaaa ctttttgaaa gctagaagag aggttaatat
8541 gaattgaanaa gtagaaggca tgaccttgga gctgaataa gaccatttag gtgtccaaga aggcataaag
8611 ggtttaaagc gacaattagg tgttgttaat agtgaatga aagctaatct gtcattattt gataagctg
8681 aaaaatcaat ggaanaagat caggcgagaa ttaagggttt aaatgataag ctttaagtc aaaaaaagat
8751 gtattctcaa gtagaagatg agcttaaaaa agttaacgct aattatcaaa aagctaaatc tagtgtaaa
8821 gatgttgaga agcctatttt aaagctagta gaagctata aaaaagabaa attagctctt gataaatcta
8891 aagaagcctt aaatcttctg aatcacagac ttaaaaaagc tgaataatca tataacgta caaatcaacg
8961 taacaagat gcatatcaaa aacttaaaaa gttgagagat gcagaacaaa agcttaagaa tagtaaccan
9031 gctactactg cacaactaaa aagagcaagt gacgcagtag aagagcagtc cgttaagcat aagacattg
9101 ttgacaataa taacaagaaa ggaactcaag ttcaaaaact aaagtagcaa aatgataatc tttcaaatc
9171 aaacgaaaaa atagaanaat cttacgttaa aactaatact aaataaagc aaacagaaa agaatattat
9241 gatttaataa atactattaa gaatcatagc gctaatgtcg caaaagctga aacagctgtt acaaaagaaa
9311 aagctgcttt aaataattta gagcgttcaa tagataaagc ttcactcgaa atgaagactt ttaacaaaga
9381 acaaatgata gctcaagtc atttcggcaa acttgctagt caagcggtat tcatgtcaaa gaattttag
9451 tctattggag ataaaaatgac ttccctagga cgtacagtag cgaaggggtt atctacacg attactttag
9521 ggttaggtgc agcttaaaa acaagtcag acttcgaag gcaaatgtct cgaatgtgag cgaattgaca
9591 agcaagcagt aagacrttaa aagcatgtc taatcaagc gttgacttag ggcataaac aagtaaaagt
9661 gctaacgaa gtcttaagg tatggaagaa ttggcagctt aggtcttaa tgcataacaa acatagggg
9731 ctatgcgggg tgttatcagt gcagcagaag caagcggtgc agaaatgggt acactgcga cgttaatggc
9801 akcagcaatt aattcttctg gtttaaaag atctgagcga aacatgttg ctgatttact tgcagatca
9871 gctaatgata gtgctgcaga tattcaatag atggagatg catttaataa tgcaggtact ccagcaaaag
9941 cattagaggt tcaatagag gacattctg cagcaattga agtttatct aactcagggg tagaggggtc
10011 tcaagcaggt actgcattaa gagcttctgt ttttaggcta gctaatccaa gtaaaagtag agctaaaggaa

5

177

10

15

20

25

30

35

40

45

50

55

10081 atgaaaaaat taggtattca ttgtctgat gctaaaggtc aatttgttgg catgggtgaa ttgattagac
10151 agttccaaga cncatgaaaa ggcataga gagaacaaaa actagcaaca gtggctacaa tagtggcac
10221 tgaagcagca agtggatttt tagcctttag tgaagcgggt ccagataaaa ttaatagcta tagcaaatca
10291 ttgagaacct ctaattggtga aagtaaaaaa gcagctgatt tgatgaagaa caacctcaaa ggtgctctgg
10361 aacaattagg tggcgtcttt gaattcgttag caattgaagt tggtaaaagt ttaacyccta tgaatagagt
10431 aggtgaggaa ggattaaaca aattagttga tggatttaca catcttctgt gttggtttag aaaggtctcg
10501 gtatgtttag cgatttttgg tgcatttatt ggcctgtctg ttcttctgg tggcttatta atacygcaag
10571 ttggaagocg ggcataaagg tatgcattcat taatagacg cattgctgaa taatcaatcac tgcataatcac
10641 caattcaaaa gcaatgaagt ctttaggtct tcaaaaccta ttcttgggt ctacaacagg aaaaacgtca
10711 aaaggtctta aaggattagc cggagctatg ttgtttaatt taataacctat aaatgttttg aaaaatctcg
10781 caaagctagc aattttaccg ttcaaacctt tgaaaaaagg tttaggatta gccgcataat ccttatttgc
10851 agtaagtggg ggcgaagat ttgctgtgtt agccttaaa ttcttaacag gacctatagg tgctacaata
10931 cctgtctatta caattgcata taaggttttt aaaaacggcat atgattcgtt ggaattggttc agaaaaggta
10991 ttaacgggtt aggaagaact ataaagtttt ttggtggcaa aattattggc ggtgctgtta ggaagctagg
11061 agagttttaa aattatcttg gaagtatagg caaaagcttc aaagaaaagt ttccaaggga tatgaagat
11131 ggttataaat ctttgagtga cgaatgacct ctgaagtag gagtcaacaa gtttaaaaga ttatgcaaaa
11201 ccatggggcag agcttctaaa aaagcattct atactgtaa agtgttgggg aaaggtgttt caaaagaac
11271 agaaaaagct ttgagaatat aogtactata ttctgaagag acaaacagaa tcatgganaa agtactgtta
11341 agactgggtc aaataacaga agcaaaagca aaaaacttt tgaaaattga agcggattta tctaataacc
11411 tcatagctga aatagaaaaa agaaaataaa aggaactcga aaaaactcaa gaacttattg ataagctag
11481 tgcgttctga gaacagaaaa agcaaaacat tttaactaga actaaagaaa aaaaatgactt gcgaattaaa
11551 aaagagcaag aactcaatca gaaaataaaa gaattgaag aaaaagcttt aagttaggtt cagatttcag
11621 aaaaatgaag aaaaagaatt gaaaagcttg aaaaataaag acgtgacatc actgttaaa aattgagtaa
11691 gactcaaaaa gagcaagagc gtatttttag aagaatgcaa agaaacagaa atgcttattc aatagacgaa
11761 ggcagcaaaag caattaaaga agcagaaaaa gcaagaaaaag caagaaaaaa agaatggagc agcaaatag
11831 aagatgatgt cattgctata aaaaataacy tcaacctttc taagtctgaa aagataaat tattagctat
11901 agtctgataa agacataagg atgaagttaa aaagcaaaa tcaaaaaag atgctgagt agacgttgtt
11971 aaaaagcaaa ataaagatat tgataaagag atggtttat caggtgtgoc tgatataaa aatactgaaa
12041 agtgggtgaa tggcctttaa agttggtgtt ctaacttcag agnagaccaa agaaagaaa gtgataagta
12111 cgctaaagaa caagaagaaa cagctogtag aaacagagaa aatataaaga aatggtttgg aatgcttgg
12181 gacggcgtaa aaactaaaac tggcgaagct tttagtaaaa tgggcagaaa gtctaactat ttggtggcgc
12251 aaatgaanaa aatgtggagt ggaatcaaa gaattccaag caaattaagt tcaggttga gctcagccaa
12321 aagtctctga ggtatcaca ctaaggttat agctaatagt actgttaaat gttttgaaa agcttggcaa
12391 tctgttaaat cgactacagg aagtatttacc aatcaaaata agcaaaagta ttcatagctc tcagataaag
12461 agtgggtgaa ttcaaaatct atttggaaag ggcactcaaa atggttttag atgcatata aaagtgcana
12531 gggctggcta acggtatagg ctaataaact ggcctcgaaa tgggataata ttcttagtac agcattggtc
12601 gactcaaat cggtttgaaa aggaacatcg aaatggttta gtaactcata caaatcttta aaaggttga
12671 ctggagatat gtattcaaga gcccaogact gttttgagc aatttcaagt tgggtatggt ctaacgtcaa
12741 agctactatt aatggtttta gaaaatggct atcaagaaca tatgaatgga ttagagatat tggtaagac
12811 atgggaagag ctggcggtga tttagttaa aatgttgcta ataaagctat tggcggttta aatagcatga
12881 ttggcggtat taataaaaa tctaaagcca ttaactgata aaatctcatc aggcataac ctacttgc
12951 tactggtact ttagcagaaa aggtgtgagc taccgataat tggggagcat taacgcaacc gacatttgc
13021 gatttaaatg atagaggttc tggaaaagcc ccaggtgggt gaggtaaaag agtaattcac aggtgtgacg
13091 gaacattcca tgcaccccaa ggcagagatg tggttgttcc actagagatt ggaataggt taataaatgc
13161 cgtgacact ctgaagtac agcggatggg tgttttgcaa aaattccatg gtagtaccga aagaagaat
13231 tggctagacc aactaaagg taatataggt aaaaagcag gagaatttgg agctacagct aaaaacagc
13301 cgcataaat caaaaaaggt gcaagaanaa tggttgaag ccagggcgat aaaaataag atggtgactc
13371 ttggttaggc gataaaatcg gcaatgtgtt ggaattagta caacatccag ggaactagt aaataaagta
13441 atgtcagggt taataattaa ttgtggaggc ggaactaacg tacagtaaaa attgctaaag ggcgtactc
13511 atgtctcaaa aagaatttaa tagacaaagt aaaaatcgtt ttggaagatt ttggttaaag aggcagtgga
13581 agctacttat tgaatatcc aatctggcaa agatttggc gctacacagg tggactaac ttcaatgag
13651 gtcgtcacta tggatagac ttgtgtatgc ctactggaac aaactttat gccgttaaag gtggtatagc
13721 agtaaaagta tggactgact acgttggcgg caattctata caaattaaag ccggtgctaa cgaatggaac
13791 tggatatagc atttatctaa gcaattagca agacaaggcc acgttattca agctgttcaa ctgataggga
13861 aatcaggtgc tacaggtaat ttcgttagag gagcacact acatttccaa ttgatgcag ggtcacatcc
13931 agggaaatgac acagctaaag atccagaaa atggttgaag tcaacttaag gtagtggcgt tcgaagtgtt
14001 tcaaggtgta ataaagctgc atctgttgg gcaaggcata tacytctgc agcaaaacga atgggtgta
14071 atgttacttc ggttgatgta ggaatatca tttagcttga tcaacacgaa tcaggaggaa atgcaggtat
14141 aactcaatct agttcgotta gagacatcaa cgttttacag ggcataccag caaaaggatt gcttcaatat
14211 atcccacaaa catttagaca ttatgctgtt agaggtcaca acaatatata tagtgggttac gatcagttat
14281 tagcgttctt caacacaga tattggtcgt cacagtttaa cccaagaggt ggttggcttc caagtgttcc
14351 aagaagatat gogaatgggt gtttgattac aaagcatcaa ctgctgaag tgggtgaagg agataaagc
14421 gagatggtta tccctttaac tagacgtaaa cgaacattc taataatga tacttctaca gttgaaaat tgtgaacaa
14491 gtaggttgg caagccaaat aaatcactg taataatga tacttctaca gttgaaaat tgtgaacaa
14561 aattgttatg ttaagtata aaggaataaa attaacagat gacttgattc aaactgttcc ttctcaggat
14631 aataacttag gtttaaatga tgcatttaga ggttagaaa aaatattgtc aaaaacaaat gggcatagag
14701 caaatgcaaa taattatag ggaaggttga ctaattaatg caatctttg taataatcat agatggttac
14771 aaggaagaag taataacaga tttaactcag cttatatttt tagatgcaag ggtgaaaagt ccaaacacca
14841 atgataacag tgtaactatt aacggagtag atggtatttt accggggcga attagttttg cgccttttcc
14911 atagttatta aggtttggtc atgatggtat agatgttata gatttaaat tatttgagca ttggtttaga
14981 tctgtgttta atgcagaca tcttattatt gttattactt ctcaaatgcc tgggtttaa atgcagtaga
15051 atacagctaa atgtttggtc atgtttttaa atgttttcc aactgaaat gaagttaagt taatgttta
15121 taaggggtat tctgaatcag ttaattggac cgaatgcag ttcttattcc acttaattg gatgtttgaa
15191 aatggaatct ctcttgattt cacacctaaa tatactcata catcaatca atttactatt tggaaoggtt
15261 ctactgtac gatcaatcca cgaattcaagc agatttgaa aatattatct aatttaaatg cgaatggagg
15331 atttgaactg gtttaactata caacaggtga tatttttaa tacaacaaa gtatagataa aaactatgat

5

178

10

15

20

25

30

35

40

45

50

55

15401 ttgtgtttag atgggtgtga tgcatacga gatataaata gagtgggaat tgatacaaat agaggcatta
15471 taacattagc gccaggttaa aatgaattta agattaagg agacatcagt gatattaaaa ctacatttaa
15541 gtttcctttt atttataggt aggtgattta atggatttatc atgatcattt atcagtaatg gattttaatg
15611 aattgatttg tgaanaattta ctatagtagt attatgggttc ttttaaggaa tattatgaac tgaatgaagc
15681 taggtacatc actttttacg tttatagaac tactcatnat agttttgttt tctgattact aatttgtgaa
15751 acctcataa tttatctatg tgaanaatac acaattatagc agacagcgcc aaagggtgaa ggtgataaag
15821 tttttattga agttacggga tatcactata tgtatgaatt tcaaaatcac tcagtggaat caaatagctt
15891 tgatgcagac agtgcgggaa ctggtaaaac gccagatcac tcttttagat agtacttaag atatggtatt
15961 gcaaatcaaa aaacttcogt caaatgacc tatataataa ttggaatttt taagcgaaaa gtaccgattg
16031 agaatatgg taacaaaaac ggcttagaat actgtanaa agcggtagac ctatttggtc gtataattta
16101 cccaaatgat aogggagatg gttttttatc tcttgaana ctttatcaaa gaagcgagaa agtgattoga
16171 tatcaatata atactgatac tgtattcgca actgtcagta cattgggaat aagaacagct ataaaagttt
16241 ttggaaaaaa gtatcacagct gaggaaaaga aaattataa tcttataga caaactgaca ttaaatattc
16311 aaatgggttt ataaaaaag gtacttatcg taccgcaaca atgggttcta aagctactat taactttgat
16381 tgcaggtatg gtaatgaanc agttagattt acaataaaaa agggctctca aggtggaata tataagttga
16451 ttttagacgg caagcaaat aagcaatttt cttgtttgc taagtgggtt cagctcgaaa caatagattt
16521 aataaaaaat attgataaag gcaagcagct tttagaatg atatttttag gagaagaccc caaaaataga
16591 attgatata cttaaaaaa aaagctcaag cttgtatgt atgttggaa tgaaaaatca acagtcttaa
16661 atttaattgc tgacaactca ggtcgcaatc aatcaaaagc aattgttgac taccgtcgag atagtgcmaa
16731 gcaagtttgg attcgatagc ctaatacaga acaaatgaa gatatcgaaa cacaggataa gctgtagaa
16801 ttgtcaaaa agcaaatata tgatactcct aagactgaat tagatgttaa ttataggtt tatgaaaaaa
16871 tagagccaag agatagcgta ttctttgttc atgaattaat gggatataac actgaattaa aggttgttaa
16941 acttgatagg tcaatccat ttgttaaaagc aatagatgaa gtgctttca gcaatgaat aaaggatag
17011 gtacaaatc aaacagcgct taacagacga gttattgcac aagataatg atataactat caagcaatc
17081 gtacaaatc ttatcacact agtactttga attctccttt cyagacaatg gatataagga gtgtattaat
17151 ataagtgcaa cagaagaag taanaatcaa gcgctacttg aaaaagataa acagctactt ccagctacac
17221 attggaagc tataaatggg atacctatg caggcagtag tgatattgat ggtttgctc aagacggtat
17291 catttcggtg gatgataaaa atanaataga taatttaaaa ataggcgag cagggaattt tcaaatagc
17361 attgtacaga aatcccaaaa cgttaattg tggaaaataa cagttgacga tagtgggaaa cttggtacag
17431 tcttatttta ttgaaagga aggtgcatta tggaaaattt gtatttaata aaggatttgg gaggtttago
17501 agtgcagat tatagagcta aggaataca aaacttaca aagatagac aatttgctt ggtttgaca
17571 acagagctta agttgcatac gaagctaaa aaactcaac acttcgcta gcaaatttat tataatggtt
17641 gatcgcaagc agcagtaaac aaactttac aaagtcaat taacgcactt gttgtggcac cagctataa
17711 cagtgctaat gagattgttc aagctcagat taatgtanaa gggaaacct ttgacactt aaagacat
17781 ttagcagatt ggaacccca aactcaaat aataaaggg aaactaaag aagattaat aagaccaac
17851 aagaattct tgatatacag tatcgttttg aacctgataa gcaagaattt ttatttgta cagaacttgc
17921 aactcttaca aatgcagtaa tgcattcctt ctggtttgat aatagaacag ccatagata catgacaaa
17991 gctagaataa atggctatat gctaagctgt ctaagaccta atggtcaatt tatagacag ccatgattg
18061 taggtggggg tcatgttaca cataacgggt atagatatat tgatgatgag ttatggtatt atagtttat
18131 cttaaatggt aataatgaga atacattagt tctgttcaa tatagccta atgtggaat tagctatgac
18201 agtatgtgta tgcagatgt atttacaaga caccagaaa aacctacat caccctgtc ataaatgaaa
18271 aagaataaa aattctatc agaatgaga gacctagaag tcaactggaa cttgaaact caatgaata
18341 tatagataa agaatgtttag acgatgtga taaaaatatt gataaagttt tgcataaat cagatccct
18411 atgagactaa caacgaaac caaccaaag cagggtgtga cttttgatga aaataactg tattgtata
18481 caggagcag taatccaaat aatagaaact atttaacggg ttctgatta gaacaggag aagaagcgt
18551 tcegggttaat gctgactatg gtggaactc agattcattt cctggcgaat ttgaggagc aagaagttg
18621 caaatatact atgcaaaaag tagtggtaa aaagctttga tctgtcggg ggtgatgaa
18691 atgaacaca tctgattttc atgattggg aaagaggtat tttagaata cttaactca gaggcgttcc
18761 ttttatcatg agtgacacag gtggtagag taactctta ccaatgagg ctgataact tagaatctt
18831 gggatgttaa cagagccagg tctttactat ttatacactg atcatacag tcaaatcag gatttccat
18901 taccagaga atggcgtgat gcaggttggg tcttggaggt taagccacca caaactggcg gtgatgaa
18971 tcagatattg acggttaata gttatgcaag gaatatgatg acttttgaaa ggttgccttc tgaagaact
19041 ggaagacttt cggactggaa ttatgtgctt aaaaatagtg gtaaatggga gagagtacct tcaatcaca
19111 caaaaatgtc agatattaac atagtaggca tctgtttta tttaactag gatgatacaa aacgtttac
19181 agattttcca actgaacgta aaggggtagc tggttggaa ttatagtag aagcttcaaa caccgttggc
19251 tttgttcata ggttagtctg taatagttgt acagcatctg ctgagatac attgaaaaat tatgatagta
19321 aaacaaagtc agggccatgg actttcacag aagggagaat tataagttaa tgagtaattt aagaatctt
19391 gtatgctaaa atttagaaaa cacagcgcat tatgaaaaa ttcaaatct agatataact tttagaacag
19461 gagagagtg ttcttctgtt ctcttttta atatcactaa aaataatcaa cgttattat tgagtgaag
19531 aaatatcaaa gcacgaatag cgaattcgagg taaggagtc atgggtgtg ctccactaga aatattgat
19601 ccattttaaag gtatttttaa atttcaatta ccaatgatg taattaaag agatggaggt tatcaagctc
19671 aagtttcggg tgcagaatta ggttaattcag acgtgttagt tctogagaga actatcacat ttaacgttga
19741 aaaaagtttg tttagcatga tccattctga acaaaaatta cacttatattg ttgaatttca ggaattagaa
19811 aaactatta tggatcgtgc gaaagcaatg gacgaggtta taaaaaatgg tgaagattat gcagcttga
19881 ttgaaaaagc taagaaaaaa ggtctatcag atattcaaat agcaaaatct tcaagtatag atgaattaaa
19951 gcaacttgct aatagcata tatctgattt ggaanaataa gcgcaagcat attcaagac attcgatgag
20021 caaagcgat atattgtaga gaaacatgaa gccttcaagc agtcagtgaa tagtgggtgt tagtcacaa
20091 gtgttctac ttcaaatgg caaaaagcta agattactaa agatgatggg aagataatgc agattactgg
20161 atttgatttt aataatccag acaaaagaa aggtgatcca acccaattta ttatgttcc gcaagctata
20231 aattatccaa gaggtgttag tactaaaggt actgtcgaa atttagtag aaotcagat tcaagcgt
20301 tgacttatcg acggaacggg acaaataaag tgtttgttaa aagaaagaa ggggttcat ggtctgagtg
20371 gtcagaatta gctattaatg attacaatc accctttgaa actgttcaa gtgcccaatc aaagctaat
20441 atggcgaata gtaacgctaa attatacga gatgacaagt ttaataaaag gtattcgggt atttttgatg
20511 gaacagcaaa tgggtgggc tctacattgt acttaaatga gatttagac caattattt tattaattt
20581 ttatgggact ttccagggtg gtgactttac agagtttggc agtcccttg gaggagaaa gatttcattg
20651 aatccctcaa atctccaga tgggtgagga aatgggtgag gtgtttatga gtttggatta actaaacta

5

179

10

15

20

25

30

35

40

45

50

55

20721 gtctacatc tttaactata tcaaacgatg tctatttcga cttagggaagt caaagaggct ctgggtgcga
20791 cgcacataga gggacaatta acsaatttat aggaagtga gaataatgca aatattagtt aacaagcgt
20861 atgagataat ttcatatcgt atcattgggt gctttgaaga aggtattgat attgaaatt taccagaaaa
20931 tttctcccaa gtttttagac cttaagcctt taatatttca aatggggaaa tagtttttaa cgaagattat
21001 tcagaaagaa aagatgacct gcatcaacag attgacagtg aagaacaaaa cactgctgct tctgatgaca
21071 tcttacgaaa aatgggtgct agtatgcaga aacagttgt tcaagtcaca angttatcga tgcaggttaa
21141 taagcaaat gcactaatgg caaaacact tgrgacacct aataaaaaat tgaagaggt taaaggagag
21211 actgaaatg cttaaatgaa ttccaccaac attcgaagat attaaacat ggratcaatt gaaagatat
21281 agtaaaagag atnctagcgt gtatgtaga: atggaagtt tagataaaga ggaatatgca attattacag
21351 gagaagagta tccagaaat cttaggtcat aggttataat cttatggctt ttaatttga ataaagtggg
21421 tgggtgaatg tttggatttc ccaacgcaga cgaacagat tggcgttaa cgcgattga agaaaatgat
21491 agactatgt ttgaaatct cagacagaa gaagacagtc tgagacgcga agaaaaatt tatgacaagt
21561 tagatagaaa ttgcgaagaa cttagggcgt acaagagaga agatgaaaa aataagaga aaaaagtcaa
21631 aaaaattaga gacatcaaga tgtggattct aggtataata gggacgattc taagtacatt tgttatagcc
21701 ttgttaaaaa ctatttttgg catttaaaag aggtgattac catgcttaag ggaattttag gatatagctt
21771 ttggctgctg tttctgttta gtaagttaa gtaattgta agagtcatgt cttcggcact ggctttttat
21841 ttggaaaaaa aggaagcaaac aatgggagtc aaaaagtaata acaagataca tctgtatgat cttgacatta
21911 gtaaatcaat tcttagcgaa caaagggtatt agcccgattc cagttagaga tgagataata tcatcaataa
21981 tcttactgt tgttgcctta tatactactg ataaagacaa tccacatctt caagaaggtc aatgggcaaa
22051 tcaaaagcta aagaatata aagctgaaaa caagtataga aagcaacag ggaagcgcct aattaaagaa
22121 gtaatgacac ctacgaatat gaacgcacaca aatgatttag ggtaggtgtt gaccaatgtt gataacaaa
22191 aacaaagcag aaaaatggtt tgataattca ttagggaagc agttcaatcc tgatttgttt tatggatttc
22261 agtgttaca ttacgcaaat atgtttttta tgatagcaac aggcgaaggg ttacaaaggt tatacgttta
22331 taatattcca ttgataata aagcaaggtt tgaataatca gggcaataaa ttaaaacta tgatagcttt
22401 ttacgcgaaa agttggacat tgtcgttttc cctgcaagt atggtggcgg agctggacat gttgaaattg
22471 ttagagcgcg taactcaaac actttcacat cgtttggcca aaattggaat ggtaaaggtt ggaacaaatg
22541 cgttgpgcaa cctggttggg gtcccgaaac cgttacaaga catgttcaat attacgatga cccaatgat
22611 tttattagat taattttccc agataaagta agtgttggag ataaagctaa aagcgttatt aagcaagcaa
22681 ctgcacaaaa gcaagcagta attaaacctt aaaaatttat gcttgtagcc ggtcatgctt ataacgatcc
22751 tggagcagta ggaacggaaa caaacgaagc cgaatttata cgttaataata taacgcacaa tatcgttaag
22821 tatttaagac atgcggtgca tgaagtgcga ttatatgggt gctcaagtc atcacaaagc atgtatcaag
22891 atacagcata cgtgtttaat gtatgtaata aaaaagatta tggcttata tgggttaaat cacagggtta
22961 tgacattgtt ctagaatac atttagacgc agcaggagaa agcgcaggtg gttggcagtt tattatctca
23031 agtcaattca atgcagatac tattgataaa agtatcaagc atgttattaa aataactta ggcacaaata
23101 gaggtgtaac accctgtaac gatttactaa atgttaacgt atcagcagaa ataaataaa attatcgctt
23171 attcgaatta ggttttatca ctataaaaa tgatatggat tggattaaaga aaactatga cttgtattct
23241 taatttaagc cgtgttgcgt tcatggtaag cctatcgggt gttgtgatac tagtgagggt aaaaacccag
23311 taanaacgca aagaatccgc ccagtgccag caggttatac acccgataaa aataatgtac cgtataaaaa
23381 agaaactggt tattacacag ttgcaactgt taaggttaat aacgttaagg ccgctatttc aactaattca
23451 agaatctact gtgtattacc taataacgca acaatcaaat atgacggcgc atattgtatc aatggctata
23521 tatttggtac tatattgctt aatagtgga aacgtcgtta tattgttaca ggaaggttag acaagggcag
23591 taataagata agcagttttg gtaagtttag tgcagtttga taattgtata tgatgaatct taggcaggta
23661 cttcgttact tgcctattat ttaaaattaa taacagctta atttttacct gnatatatta aattttaaaa
23731 caaacaaactg ttttagcata taattatttt tgtgttctga tttgtgtgca tgatcaaaaa gttgttatgg
23801 tcaactatat cgtgttttta tgtttattat caatcaaaat ataaattatt tataatttgt ttgtaatatg
23871 acgggttttt ttgcgaataa tagtaaaaaa acacatttgt agatatttta aactcggtaa atcttttaat
23941 aaattattaa ttctattaaa agttaaaaag gtttaataa aaaaatgaat aaaaattata aagaagggaa
24011 atgattttta tggtaaaaaa aagactatta gctgcaaat tgtcgttagg aataactact cctattgcta
24081 cttcgtttta tgaattctaaa gctgataaca atattgagaa cattgggtgt ggcgtgaggt tagtcaaaag
24151 aacagaagat acaagtagcy ataagtgggg ggtcacacaa aatattcagt ttgattttgt taagataaaa
24221 aagataaaca aagacgcttt gattttaaaa atgcgaaggt ttatcaattc aagactact tattacaatt
24291 acaaaaacac agatcatata aaagcaatga ggtggccttt caatacaat attggttca aacaaatga
24361 ccccaattga gatttaataa attatctacc taataataaa atagattcag taatgttag tcaaacatta
24431 ggttataaca taggtggtaa ttttaattgt ggtccatcaa caggaggtaa tgggtcattt aattattcaa
24501 aaacaattag ttataataaa ataaaaagta ggtgataaga tgaactcaat tctagggcgt cttcttctta
24571 caggagtttt aggttncatc ccatataaat atctaacaat gataggttta gttagtga aaacaaggt
24641 tatcaattact cctgtattat tgattttttc taatgaacaa gtttgatatt ggttttagg ttttatatt
24711 tttaataatg ttgattttaa aaatttgaat ttaattcagt tgcctacagg tctaaagca aatattttgt
24781 ttctatttat ttttgttta acagtgtttg ttttaattcc ttaatttgtt aaatttatta tctggttaat
24851 taatataacc agaaagttaa tgaatttga ttgtataagc ttattagaca aaagagacaa gttgttaaat
24921 aacaacgcta aaccagttat tatagttata aaagactttg aaacagaaat cattgaagag ggtgaactta
24991 aaacctataa ttacgtcgtt agcgaatttc atttactaga agttgagcga aggaatttca aagtatctga
25061 ttacgcgta aacgatgaat tgtatattaa acatacact gttagacctt aacaacaaat taatttggat
25131 ttatatttaa tgaatgaata ctaattcttt tcttagottt tttctgataa agtgcctttt aattttctgc
25201 tggcggccgg cttttcaaaa cttttgttta ttgggttact acgagtactt tcttgttttt tgtttttatc
25271 cgcataaaa ttctaccac cattcaacgt ctacacttgt aggcgttttt ttatttagta aagtcataat
25341 gaattctctt tggtaactt atctccatct attttttgtg aaataaatc caagtattta cgcgatttat
25411 gtgacgataa atcttttagt aactcataag tgaatgggtg attaccacta gttaaaactt catatactat
25481 agtttctttt ttatttttgc aattagttat ttctattata aacttctttt caaacactgc tgaatagac
25551 gctttttata ttaagcgcac acacaggcgc tgttaatac aatacaactt tgcctattac ttttaattta
25621 ctaaaagcag cgaactttgat atcatcatac ttoggattta gagataccaa attaatatag tcttcgcata
25691 tatctacacg cttgataaga cttaactccat ctaatacaac gaggtaaat gtaccatctt taatagatc
25761 ttctttctta ataaaagcgt atgttctctg ttttaacata ggttccattg aatcaccatt aactaaaaa
25831 caaaaactcag catttgatgg cgtttcgtct tctttaaaaa atacttcttc atgcaatag tcatcataa
25901 attctctctc tatgcagcga ccagttgtac cacatgcaat atacgatact agtttagact ctttatctcc
25971 atcatagaa gtgactttat tctgtttctc caattgttca ttgtcatagt taagtacgtt tttctggcgg

5

180

10

15

20

25

30

35

40

45

50

55

26041 ggagggtgtga gttgtgtgtga tatggaaagt atgtcgttat cgtctttgtga cgtagatttt gattcactat
26111 acaaatcatt aatcttcaca ttgaagtact cagccaaaat ttgtggcagtt gataatcgag gttcttccttt
26181 ttcatcttcc catttttgata tcttgctttt cgttaatttc attaatgctgg gatattttatt attaatgata
26251 gttgttaatt gttccatagt catattttta tttttttctt agcttcttta aaccttcacc aataccata
26321 cgaacccctc ctatataang ataattttcat taaaaagtt cgaacacga aacgcaagg aattttatt
26391 gcaaaagtgt ttgacatoga aacttttatg atgtattctt aaatcaagtt gttcaaacg aaacaaaagg
26461 aggggggtta atgacaacta gtgtagcaga taaccatac ttaaaaaaa aaagcttgat tgcacttaaa
26531 ggaactaaccc aaaaagaagt tgcataagca atcggaatga gtagaagttt attgagtata aagataaatc
26601 gaattaatgt cagagatttt acancttcag aagctaaaa attagcagat catttaagt ttaagttga
26671 tgattttttt taaactttta gtttcgaag tgacaactaa ataaaaaa ggaggacact atggaaacaa
26741 taagtttaac caaagaagag ttgaagaaaa ttatagcgaa agaagttaga aatgctataa aaggcgagaa
26811 accaatcagc tcaggttgcaa ttttcagtaa agtaagaatc aaatagcag atttagaaga aatcaatana
26881 aaactcaatt tgcgaagaa ttgttcgcta ggaagattga ggaagctcaa tcatccgatt ccgctaaaaa
26951 agtatcagca tggcttcgaa tcaattcacc aaaaagctta tgtacaagat gttcatgacc atattagaaa
27021 attaacatta tcaatttttg gaggacact taattcagac ttgagtgaaa gtgaatcaca cctagcagca
27091 aaattttata gatatacaca aaactattat ttatctatct atgaaaagag agtttcagaa ttaactatcg
27161 atgatttoga atgaaggagg aactcaaat gaactacta agaaagctat tcaataaaaa acacgaanaa
27231 tcaattgagc tgggttcgag aaactaatgg ttaaaagta agaaagcaca ataaaaaa tatagagtg
27301 tctogtagag agaaggttaag aaactctaa ttaaaagca gcaacttaatt agtgcagta atcaagtcg
27371 ctattgcctt acaactcctaa atctttctg ctcttttctt ctctctgtta tcccaataac acgaagag
27441 aaagctgtaa atagtacga gcaacgtat cttagogaa tgaactacg tcatcacoga ctctctgcaa
27511 ttogttatga atcttatgtc tatctagagc tctaggtaat agcgagattg taatattctg agcaattttc
27581 tctaaatcca taaatttcac ctcttccac tgggagataa cttaattata taacaaaaca acttaaggaa
27651 ggaacgacaa atgcaagctc aaaaacaaaa agtcatctat tactactatg acgaagagg taataggcga
27721 cctatagata ttcaaatcaa tgacggatag gaactgatgg tccgattcoa tttcatcaac aacaccattg
27791 aagaataacc atagtaaat aaactattat atgctctggg ttaggtttat gaatttaagt tagattgaat
27861 ttttgagaaa gatattgaaa agtcaatttc cccataagat taagagacat actggatgtt tgttaacga
27931 ctcttttaac ttctgtccaa gttttattgt ctctaatatt atcgagaat tcatggccag accaagtgat
28001 tctaatcaata atccaaagaa cgacctgccc ttcgatgaat ttcagatcgc acaaatanaa tttagctct
28071 tctaatttta aaagttagta cactactgtt tcaaaatcat atttatcaa aataattata tctgtgaat
28141 tatctogagt aagtggttca cctattttct tattagattc tatttctaa agcaagagtc taacgaatc
28211 gtagtaagt ttcatcctat cactctcaca acaggagat agcagaagg atcataaca tcttaaaagg
28281 aggaataaca aatgaacatt caagaagcaa cttaagatgc tacaanaaat ctgtctcta tgacacgaa
28351 agattgaaa gaaagtcatc gaactaagt attaccaaa aatgatagtt tttaacatg catctttca
28421 aatgcgatg ggaacaaact tatcagatat tggcaacctt cagcogata cctcatggca aatgattggg
28491 aagttataaa ccaactaga gaccaggaat tattgaagca attttgaaa tgcatacat gatattttt
28561 aaattgtttt taaactcatt ttcaagtaa acaacagctc tgtctgaat tgttactatg taatagtg
28631 tactagcata cagcgcgttt aggaacccag agtttttaag ttattttaa tctatttta catcttcga
28701 taattgtttt tgaaaacat ttgtatgtat atcttttagca ctccaaaat tattgcaggt taatttaacc
28771 gaacttaact ttacacattc taataatctt ttgtagatga cgaacaaat atattgttg tctttagtaa
28841 gttgatcaaa ttcatcagat atcaaggcca ttttatcacc tcottaggtt gatacaaca ttatcacga
28911 aaggagcata acaaatgaa cacaagatca gaaggatgc gtataggcgt ccccaagtt tctagcaag
28981 ctgattgttc ttcatcctat ttaacggaaa aggaagctaa cttaggagcg gaattattag agcttataa
29051 aaaaagtatg tacagctact tagaataaaa caaagtttc tatgcttag atagagaact tcaatcagg
29121 gogaaataa acaacttta acattttatc aaaggagtg tagagatgcc aaaaatcata ataccacca
29191 caccagaaa cactatcga ggcgaagaaa aatttgttaa aaagttatac gcaaacctca cacaatcca
29261 tcaattgttt ggaatattga gaagtacat atacaactgg ttgaatat accgtgaaga taatttaggt
29331 gtgaaanaat tatcactga ttattcagca acgggaactc tgattaatat ttctaaata gaagagatt
29401 tgatcagaaa gcatanaaaa tgggtattagg aggtattca aatgagcag acatataaaa gctacctatt
29471 agcagtggtg tcttcacgg tcttagcag ttactcatg ccgtttctat acttactac agcatggtca
29541 actgcggat tgcgaagat cgaacattc atattttata aggaactatt ttatgaaga taaganaact
29611 gctactggtt ggaagagta acagtcaag argagcaatt gcttaataa attatataag gatttataa
29681 tatgacctta caaanaaaa tactatcaca tttgcacaa tatgacaatt tcaattctga ttaggtttt
29751 gaagtttttg ggaatctaaa aacacatgca aaatccacac ttcaagact taagaaaaa ggaagattg
29821 aattgaaaag ttgggttctc tgggtgtgtg ttgaacgca gttacattt actgtgttag aacgtaaaga
29891 agagatacta gaagaacaa tgcagttatt ggaagatta aagcaaaaa gttatagacc tagaanaa
29961 gaagaacgca tcaagttat gatctgtta gccaacaaat tttaaggagg agttaatcaa tggcaatatt
30031 agaaagtatt ttgaagaat taaaactatt aaataagaa ttactgtgc taataatga actatcaat
30101 gtgatttcat caattgtaca agagaagtt aaagaagcac caatgcacaa agatgaaca gctcaactg
30171 aatcagttga agaaagtta gaaactctg ctgatttaac taagattat gttttatcag taggaanaa
30241 gttctttaa aaagcagata ctctgtataa gaaagattt agaaataaac ttaacgaact tgggtcggat
30311 aagctatctc ctatcaaga agagcattat gaaaaaattg ttgattttat gaatgcgaga ataatgcat
30381 gaagctagat cactcaaaa gagctcatgc aaagcttag gcaagtgag caaaacaaat gctaaactg
30451 ccacagagta ttaaggcaag tgaagttatt gcagataaaa gttcagttt tgcgtgaaga ggtacattc
30521 ctcatgagt aagtgaatta tatttcagtc ttaaatatga aggcctaaca cagtttgagt ttaataaag
30591 ttttcaaat tataagcga atcaatatta cagtgaagag ttgcggaat atgttgaga gtaactagct
30661 aatgtagaag aaaaatataa cgaagcttg agtagagag acgatgaat agctttattt gaaacaaat
30731 tggatttagg taataacgtc cctgaatctt ttggtactg ttggtactt atattttcag ggggtgact
30801 tgaattatt gatcttaaat acgttaagg cattgaagt ttagctatag ataatctca acttugata
30871 tatggcttg ggcataatga actgcttag ttaattgatg acattcatc agtttcagat actatcatc
30941 aaccacgaat agataacttt tctactgaag agttaccaat atcaagatta cttaactggg gaacagatt
31011 tgttaacca tttagccagac ttgcttataa cggtaagggt gattttaaag caggtagtc tbttagattc
31081 tgtaagataa agcattcatg tagaacagt gcagaataca tgcaaaatgt ccctcaaaag ccacacatt
31151 ttgttagtga tgaagagatt gcagaacttt tatataaact gcttgacatc aaaaatggg ctgatgaagt
31221 agaaaatat gactagatc aagcgaaga aaatgataaa aactattctg gttggaagct tgaagaagt
31291 cgttcgcaa gaatgataac tgatacaat gcaacgttg aaagttagt tgaagcaggt tataaacctg

5

181

10

15

20

25

30

35

40

45

50

55

31361 aagatattac agaaaccaag ttacttagca ttacgaattt agaaaaatta atcggaacaa aagcattttc
 31431 taanaattgca gaaggtctta tagaaaagcc acaggtataa tcaacacttg ctaccgagtc tgataaacga
 31501 ccagctataa agcaattctg tgaagatgat ttgacaaac tataaaattt aaaaaggagc gtatataaac
 31571 atgaagacaa aagratataa taanaactaa gtgattacag gaaaagtaag agcatcatat gcacatattt
 31641 ttgaacctca cagtatgcaa gaagggcaag aagcaaatga ttaactcagt ttaactcatt cttaactaga
 31711 tacaagctca ataaaagcca ttgaacaaag tatagaagct gctaaagaa aaggaataat tagtaagttt
 31781 ggaagcaag ttctcgcaaa tctgaacctt ccattacttg atggagatag tgaagagaa gatgatgtga
 31851 attatcaaga cgtttatttt attaacgcat caagcaaaac agcactcgtt attattgacc aaaaacaaat
 31921 tagattcaag gattctgaaa ctattgtaag tgggtactat attagagctt caatcaattt atttccattc
 31991 aacacaaatg gtaataaagg tatcgagctt ggaatgaaca acattcaact tgaagaaaaa ggcgaacctc
 32061 ttggcggtgc aagtgacgca gaagatgatt tcttaagaaa attgaggtgc caagaaattg aaatttatga
 32131 caataggtgg ggttttttag cccactttta atatttcgaa atgtggtgc tataaataca cagaagctga
 32201 atatagatat tgaacatat agcagtaacg atatttcgaa atgtggtgc tataaataca cagaagctga
 32271 agatttcgaa atcttaatta tagcttattc aatagatggg ggaagcaata atgtggtgc tataaataca
 32341 gtatataatg agcttttcca cgtgatttat gagaogttta aaatgctctt atttgacctt gctgtaaaaa
 32411 agtatgctat caatgctaat ttcgaagaaa ctgtctctgc taacactttt cgttgatata agtgtgagaa
 32481 aagatggatt tgcacaaagg ttaattcaat gctgattggc atactgctt cgttgatata agtgtgagaa
 32551 gtttttagac tacaaaacca aaaaagataa gcaggtataa atttaacttg tttattctct atactgata
 32621 agccaacaaa agttnatgga ggaagaacaa gaaatttgcc tgaacatgat ctgaaaaaat ggcacactt
 32691 tatagattac tgaattcgag atgtagaagg agaaatgaca attgctaata aaatttaaga ctttccgata
 32761 actgtaattg aacaagcata ttgggttttt gacaaacata taacagacag aggtatttaag ctttctaat
 32831 catgtatgtt aggaagctaat gtgctogata agcagagtaa agaaagattg cttaaacaaag cttaaacat
 32901 aacaggttta gaaatccta atagctctac acagttattg gcttggttaa aggtatgaac aggtattgat
 32971 atactcaatt tacaaaagaa aacgggtcag gattacttaa aagtagcaac aggaagagct aaaaataatg
 33041 tagcataatg attgcaaatg tctaaaacca gtgtgaaaaa atacacaaa atgcatgaca tgatgtcag
 33111 tgatgaacgg gtaagaggtc tggttcaatt ctacgggtcc ggtactgaaa gatgggcagg tagaggtgta
 33181 caacttcaga atttaacaaa gcattatatt tcaatagctg aattagaat agcaagagat cttattaaag
 33251 aaaaacgttt tgaagattta gatttatatt tcaatgltca tctcagaag ttatttaagt aattagttag
 33321 gtagacattt actgctgaag aggttaatga actagcagta agtgattttt ctgcaataga ggcagagctc
 33391 agagcatggt atgcacaaaga acaatggcgt tttagatgtt tcaacacaca cggaaagata tatgaagcat
 33461 cggcttctca atgttttaatt gtaccggtag aaagcataac taaggggagc cttctcagac aaaaaggaaa
 33531 aggtctcgaa tttagcttag gctatcaagg ttggcgctga gcttttaaa caatgggtgc attggaatg
 33601 ggcattgaa agaacagattt acaaggttta gttgatgtt ggcgtaacgc aaatcctaac atagtttaatt
 33671 ttggaaaggg ttgccaagag gctgcaatta atactgaaa atcccgaaag agcatcata cactggagct
 33741 tagatttat atgaaaaaag gttttctaat gattgaactg cctagtgaaa gagcttttag tttccaaa
 33811 gcttttagtt gtgaaaatag ttgggttagt caagtgttg aattttatgg gtttagatctt aacgttaatt
 33881 gtcgaagttt aaaaacgtat ggtgggaagt tagtcgaaa tattgttcaa gcaactgcaa gggatttaatt
 33951 tgcgatttct atagcaagggc ttgaagcatt aggttttaaa atagttggcc atgtccatga tgaagtaatt
 34021 gtgaataaac cttagggttc aatattatgc aggttaaatc aaggaatcgc aaactatcat gaataagcct gttgatggg
 34091 caaagagatt gaatttgaat agtgacgggt ttaactctcc gttttatag aaggtatagg agtgatgtg
 34161 catgcaaatc caagctcata tcaatgcttc gttgacatt agaatccta cagaagtcga agtggttaatt
 34231 taataaacgt tatataatc aggttttag atttgaaggt ggaatgatgg cagaataata tcaatattct
 34301 taataaacgt tatataatc aggttttag atttgaaggt ggaatgatgg cagaataata tcaatattct
 34371 gtcgatttca aaggaanaact aatgacattg agagaagttt cagaataata tcaatattct
 34441 tttagattatg atacaaacat agagtcaaat aagagatgaa gaaagagaga gogaanaaat cagaanaaaa
 34511 agatgaattg aatataatgc aggtcaaatc aagagatgaa gaaagagaga gogaanaaat cagaanaaaa
 34581 gogattttga acctatacca acgnaatgct agagcggat atgaagaaga aagagagaga agattgagac
 34651 catggcttta tgaaggaaag agcatatag agcgtatca gtaacagaaa agtagatatg aacgaagcgc
 34721 aatgttcaag aatggggtg agcatatag agcgtatca gtaacagaaa agtagatatg aacgaagcgc
 34791 aagacaatgt taagcaacca ggcactaca catacgcca catgaaatt atagatttta tgcgaagcct
 34861 taaggacag tatccacttc aactagcatt cgcactaggt aatgcaataa aatcttctgc tagagcactt
 34931 ttaagaaatg gtcatgagga tttagcaaa atacttaaac caattttctg gatctaaag atactctgat
 35001 gatgacattg acagtagagg catgtaaga atacttaaac caattttctg gatctaaag atactctgat
 35071 caggatcaag aacgagtggt acatatccat atacgggga agagctcgaa acatatataa agcaacatgg
 35141 caggctggca agcggtgaaa aagacatttg atacgggga agagctcgaa acatatataa agcaacatgg
 35211 ttggaatag gaggaaacga agcaactaac tttattttaa ggaatagaaa atgtagaaa tcaagttga
 35281 aaaaataatg aaaaatagag aataacttaa gtggcgcgga gaaatccgg agctatcatt tggcagaaaa
 35351 tattatacaa cagacaaaaa tgaatgaaac tttatttact tgaagaaa gtaacggag aactaagttt
 35421 gogattttat attagtttaatt gctactttta gtgtcaaaat tgaagaaa gtaacggag aactaagttt
 35491 tgaatagttt tttgaagttg acgagattca agaaagagtc tataaactcg catctataa gaaatgtaatt
 35561 ataaacgaac gtttaaaaaa tgcagaat tttcttgcta aagcatttca catcttaaac gacgacttaa
 35631 ctatgacgtt aatttggaag gaaggagagtt gatttaata atggaacag gttcaaaaa atactacga
 35701 aagcaagtg aatctggtt tgaatgaaga agcaagtttt tgaagcaag tgaatgagct attggagata
 35771 tagctaaagt aagagagtg acacaagagc tggagaaaga agcaagtgca tgggaataggt attgagag
 35841 cgttgaaaaa gatttaataa acaatatttg aatgaacag gaaagagctt aatttggaat ggaattaaac
 35911 aataaaattt ttatggagga agacgaat gaaatccgc gaaatccgc gaaatccgc gaaatccgc
 35981 agcgggtata acggcaatga cagagggga ttattaaaag aggttggga cgtgtataa aagcggcag
 36051 cgtctgatga aatcttgag ggtttacct atgctatgca agotgcaact aagaaagata ttggtctga
 36121 tgaagcagta ggaattatga cgggtcaagt tgcctataa tatggaggag agcagaaaaa tgactaacat
 36191 attcaagtg aaactattat caaagacgc tagaatgcca gaacgaatc atagacgga tgcaggttat
 36261 gacatattt cagctaaaac tgcgtactt gacccaaag aaaaagcag gatcaaaa gatgtagctg
 36331 taagcattcc agagggtat gtcggttat taactagccg tagtgggtga agtagtaaaa cgcatttagt
 36401 gattgaacaa ggaagatag acggggata tcatggtta ttagggatta atactcaaga tgaatgaag
 36471 acgttagaga gtgagatat gaggtaattt ggtcggagtc cttctgggt agatggaaaa tacacctc
 36541 taactgtaac agataaattt ttatgtatga atgtgtta tgcataaat aagggcaga aactagctca
 36611 attggttacc gtgctatat ggaacactga actaaagcaa tggaggaatt tgaagaggtt ttcagaacgt

36681. ggagcaaaag gcttcggaag tagcggagtg taaagacata ttagatcgag tcaaggaggt tttggggaaag
36751. tgagtgacat gttagaataa tttttcatag ggtttgggtg tttcttattt tgcgcataag gtatttattt
36821. tctcaagngt aaaaagacta tacacacaaa cctatatgaa atgttggtga ttgctactat ctttgtgaca
36891. tctacatttg ctgataaaca tcaaaagagc catatcttaa tagcattttt agtaatgttt tttatgagta
36961. agctcaaaac agttcaaggg agctatgagg aatgocacaa taccctagta caacatttaa agattcaaca
37031. ggaagtaagc atacacacat aactaaagct aagagcaatc aaaggtttac agttgttgat gcggagagta
37101. aagaagaagc gaaagagaag tagcaggacac aagttaaaaa aaatgcagtt attaaattag ggcagttgtt
37171. tgaataataa agggagtggt ggaatgact aaacaaatag taagattatt attcttacta gcgagtgtatg
37241. agctaggcaa gatgttaact gagcaagtat atattatgat gacggctaat gatgatgcag aggcgcgcag
37311. tgactttgaa aaactcagag ctgaagtttc atggtaatag ctattatcat ttttgaatta attatattaa
37381. tgcgttttagc aatagcactg gagggtgtgt aaatatgtgg attgtcattt caattgtttt attctatttt
37451. ctattgtact tgttaagtag catcttccat aagatgaaaa ccatagaagc attggagtat atgaattgctt
37521. attctttcaa gcagcttgta aaaaataatg gtgttgaaag tatagaagat tatgaaatg aagttgaagc
37591. aactagaaaa agatttanaa gcttaagaga ggcgttggtt tctctgtctt atttanaata atgaaaggag
37661. ccgaaacagt tagacaaagt cactcaaatg gaacaaatta aatatgatgc tgatgtttca tttctttatg
37731. ctgctagtgc tttatctaca cattggacta atcacaatat ggccttggtt gactttatgc agaagctagc
37801. acaaacagtt agaaactaaag aagattttaa tgagtcaaat aaatgttcta agtctgaaca agccgatata
37871. aaagatgttg gcggatttgt cgtgtgttat ttaaaagaag gcaaacgagc tgcgtgtcaa gtcatgaatc
37941. gtcaatgtt aacacttgat atcgatttat ctgctcaaga tatgactgac attatttcta tgttttatga
38011. tttttcatat tgtttatatt caacacataa gcataagag ataatgtcaa gactgcgttt agtattctt
38081. ttaaaacgaa agttaaagct agatgagtat gaagctattg ggcgttaaat cgcagatalc gtggcatgg
38151. attacttcga tgatacaact tatcaaccac ataggttant gtaattgocct tcaactagta acgatgcgga
38221. attttttttt ccttatgaag atttaccctt gtagagccca gataaaatat taaatgaata tgttgattgg
38291. actgacacat tagaattggc aactgtttca aggaagagga gtaagactaa aagattagca gataagcag
38361. gcgacccaga agaaaagcgc ggaattgttg gtagcatttg tagagcctat acgatagaag aagctataga
38431. aaacttttatt cctgatttat aogaaaaaca tttctactac cgttatacct attcatgaag ttcacttga
38501. ggtggatttg ttttatcaga aaatacaag tttgcttatt ctcatcaaa tacggatccc taaagcggta
38571. tgcttgtaaa cagttttgat ttatgaagca taccattata tgggtgtcaa gatgaagag ctaaaacaga
38641. taactccggtt aatcgactac ctagtataa agcaatgcag caaagagcgc aaatgatga agttgttaaa
38711. aagcaattaa ttaaacacaa aatgttcgat gcaatgcagg atttcgatga aatagtaaat agcgtatgat
38781. ctactccggtt aatcgactac attacttga aaggtacttt caaagctagt atcccaata cagaatttat
38851. attgogtaat gatccaaatt taaagggaaa aatagcattt aatgaattta caaaacaaat tgaatgccta
38921. gggaaaatgc catggaataa taatttttaa atagctcaat ggcagaagc tgatgatagc agtttaagaa
38991. gttatatcga aaagatttat gatcatcacc attcaggcaa acaaaaagat gccattataa ggcgtgaat
39061. gcaaaatgcc tatcatccag taagagatta tctaataaa atactgtggg atggacataa acgtcttga
39131. aagttattta tcaaatattt aggtgttga gacactgaag tgaatgaac aactacocaa aagccattga
39201. ctgctggaat cgtctgagta atggagccag gatgttaatt tgactatag ottacacttt atggtcttca
39271. aggtgtaggt aatctgctt tgctaaaaaa aataggtgtt gcatgttttt ctgaactgtt agtttctgtt
39341. actgtaagc aagctataga ggcattacaa ggcgttttgt taatggaaat ggcagaactt gcagctacaa
39411. gaaaagctga agttgaagct attaagcatt tcatatctaa acatgttagc cgttttctgt ttgcttatga
39481. acattatatt gaagattttt caagggcaatg tattttcatt ggtacaacta ataaagtga tttcttaaga
39551. gatgaactgc gtggaagcgt tttttggcca atgactgcaa atccagagag agttgaagtg aactgttcta
39621. aactaaccaa agaaagagatc gaccaaatct ggcagagaagc taaatactat tatgaacaa gagaagagtt
39691. gttcttcaac cctgaactag aagaagaatt gcttcaact caaagttaac atactgagga atctccatat
39761. acaggtatta ttgatgaata tcttaacacg ccaatcccaa gcaattggga agacttaact attttgaaa
39831. gaagagcatt ttatcaaggt gatgttgata tgttaccaa aggaattgta gattacattg aaagagacaa
39901. ggtctg-gcg cttgaagtg ttgttgnatg ttttgtaaa gataagggag atagttaggg atctatgaa
39971. attagaaga ttctcaagct cttgaacaa ttagaacatt ggtctgtata tgaagggcaat aanaagtgga
40041. aaattcgatt tggaaaagat tatggtgtac agatagccta tgaagaagat gaaagtttag aggtattcat
40111. atagaataa ttgaataaat atacattttt agatgttga tcaaatgttg catcattttt tgagtcatgc
40181. aacacgggtg tgaataaagt aatcgtaggt gttgtatcat ttttggtgat gcaacattga tgcaacaaat
40251. gatcacacac cttcttccct tctcgtcta aggttcaacc ctggtttgtt ccaatgttg atcaaatcca
40321. ctataaagtt taaaagtag ttttagggag taaaggggta taggggttaac cctctaagc ctatttttaa
40391. aagtttggca gaatttgat caactcga acacaaatat aaattttga tacaagtgta ataaatga
40461. gatcgacat tagaataata tttagtga gagataacaa agttaatgg attatgttta aatgggtgc
40531. caactggaaac aagaggtgta ccagataga ttattattat gccagaagga aaacattatt ttgtagaat
40601. gaagcaagaa aggggaagtt tacatcttct acaaaaatat gtgcacggc aatttgaaaa cagagatcat
40671. acagtgtagt tggattgaa taaagaacaa gtaaatattt ttataaagat ggtaggtgga acatttggcg
40741. attgatttca aaccacatag ctatcaaaag tatgcaatag ataaagtgat tgataatgag aaatcgggtt
40811. tgtttttaga tatggggcta gggaaaacag tatcaacact tacagcattt agtgaattgc agttgttaga
40881. cactaaaaaa atgttagtca tagcaactaa acaagttgct aaagatacat ggggttagga agttgataag
40951. tggaaacatt taaatcatct gaaagtgctt ttagtcttag gaacaccta aagaagaat gatgcattaa
41021. acacagaggg tgatctctat gtaaccaata aagaaaaatc taaatgggta tctgatcaat ataaaaaga
41091. atggccattt gactgggtg taattgatga actgtctaca ttcaaaagtc ctaagagtca aggttttaa
41161. tctattaaaa agaaattacc actcataat agatttatag gattaaacag aacacctagt ccaaatagtt
41231. tcaaggtttt atgggctcaa gttatttga tagacagagg cgaagaactt gattctcat tcaagcttga
41301. tctgaagaaa agatattatg caacacatca agttagcga catgttttta actgggagct aagagaogga
41371. tgctgcagag agttgactat aaacaaacag tagtcttctc tgaanaagaa agaaagtat atgaagaatt
41441. agaaaaaac tatattttag taacgggca gtttatagag atgttgaaag tgaagactt atacatgata
41511. caaaaactac ttcaactatc taacgggca gtttatagag atgttgaaag tgaagactt atacatgata
41581. agaaagttag taagttagag gaaattatag aggaagctca aggcacaca atattattgt tttatcaatt
41651. caaacatgat aagaagaaga taactcaagc gtttaaggaa gcaacacat tagagggattc aaactataa
41721. gaaagttgga atagtgaga ctttaagctg cttatagcac atccagcaag tgcagggcat ggaattaaat
41791. tacaacaggg tgggacatt attgtttgtt ttgacttac atggtcattg gaattatacc aacaagcaa
41861. tgcaagatta tatagacaag gaacaaatac taagactatt attcatcaca taatgaccga taacacata

183

5

10

15

20

25

30

35

40

45

50

55

```
42001 gatcaagag tatatnaagc ttacaaaat aaagaactaa cgaagaaga attgatgaaa gctattaaag
42071 caagaatagc taagcataag taatggaggc ataagatggg aaaggcgtca tatgatatta agccaggaac
42141 atttaaatat attgaatcag aatatatataa tttaaatgag aacaagaaag agataaatag attgagaatg
42211 gagataactta acccaacgaa agaaactagac acccaacttg tcatgggacc gttacaaaag ggagagccag
42281 ttagaacaac tgagttaatg gcgacaaggt tattgactaa taagatgta cgttaacttag aagagatggg
42351 tgagcaggtt gaagtgagt acttaagtt acctgaagat cataagaag taataagggt aaagtattgg
42421 aataagcta gaagctaaa gatagaacaa ataggggatg cttgtccat gcatcgcaat acagtacta
42491 caatacgaaa gaactttgtt aaagcgatag cgtatcatgc aggtatcaaa taacatttg caagattgt
42561 gcaaaaggcc tacaaatctg tagtaatatg atagtatcgg aaagatgta aaagtattct gaaagttata
42631 cgacataaat acatgaggca catcgctaag cggctgtct tttgttatgc aatcaaaag gtgtaagaga
42701 tgaccaagca taataacatt tataagcatg gtcgtaagtc atatcaatc gattggctct atcattcaaa
42771 agcatggag aagttaagag agatagcatt agatagagat aattatcttt gtcaaatgtg ttacgcgaa
42841 gatattataa cagatgcaaa gattgtgcat cacattattt atgttgatga agattttaac aaagctttag
42911 acttagataa tctaattgca gtttgttata gctgtcatan cnaaattcat gcaaatgata atgacaaaag
42981 taattcttaag aaaattagag ttctaaaaa: ttaaaatana aattattta aacaaaattt tatgcccccc
43051 tgcccatcgg cttaaaatgt tttttcggc ggtaaccgag aggc
```

Table 8

Bacteriophage 3A ORFs list

SID	LAN	FRA	POS	a.a.	RSS sequence	STA	STO
100379	3AORF001	1	8515..13488	1657	acaggtacggatttaagaaaacttt	ttg	taa
100380	3AORF002	2	37667..40114	815	tttaaaataaatgaaagagccgaac	atg	taa
100381	3AORF003	1	32188..34149	653	tttaaaagaatttgaggtgtcaagaat	ctg	cag
100382	3AORF004	3	17457..19370	637	gctattttattagaaaggaaggtgc	att	taa
100383	3AORF005	1	334..2034	566	agaaaaagatagttcaagaagaag	gtg	taa
100384	3AORF006	1	15571..17154	527	ctttttattataggtaggtgattta	atg	taa
100385	3AORF007	2	19337..20836	499	atgatatataaaacaagttcagggcc	atg	taa
100386	3AORF008	3	22176..23630	484	aatgatttagggtaggtgttgacca	atg	tga
100387	3AORF009	1	40726..42093	455	gtataactttttataagatggtag	gtg	taa
100388	3AORF010	3	13491..14738	415	gagcggaactaacgctacagtaaaa	att	taa
100389	3AORF011	2	2039..3277	412	attaaagacataatgcgttaaggag	gtg	taa
100390	3AORF012	2	4001..5209	402	aaaaaagagaaaaaattcaaacgcga	atg	taa
100391	3AORF013	1	30379..31545	388	attttatgaatgcggaataaatgc	atg	taa
100392	3AORF014	2	14738..15562	274	atttatatggaggtttgactaatta	atg	tga
100393	3AORF015	3	3249..4034	261	cttgatttaagaanaattctttgaag	gtg	tga
100394	3AORF016	-2	25587..26273	228	aagaaagctaaagaaanaataaaaat	atg	tga
100395	3AORF017	3	6729..7370	213	tttaatttttaaggaggaataaagca	atg	tga
100396	3AORF018	3	24540..25154	204	aataaataaaaaagtaggtgataag	atg	tga
100397	3AORF019	2	31565..32128	187	ctataaaaatttaaaagagcggtat	ata	tga
100398	3AORF020	3	36150..36713	187	gcagtaggaatttatgcgggtcaag	ttg	tga
100399	3AORF021	2	24011..24535	174	gtataaaaatttataaagaaagga	atg	tga
100400	3AORF022	-2	12423..12938	171	taaagtaccagtagacaatgtaggt	att	tga
100401	3AORF023	1	7462..7917	151	aaataaatacaaggagaataatttt	atg	tga
100402	3AORF024	1	26731..27174	147	actaaataaaaaataaaggagacact	atg	tga
100403	3AORF025	1	42106..42543	145	taagcataaagtaaatggaggtataag	atg	tga
100404	3AORF026	2	35255..35671	138	aagcaactaactttatttttaaggag	ata	tga
100405	3AORF027	2	5888..6298	136	atattggctataaatacaggtgtttt	acc	tga
100406	3AORF028	-3	27845..28255	136	cttttttaagatgtttatgcattttt	ctg	tga
100407	3AORF029	3	34344..34748	134	tttaaggttttagatttagaggtgga	atg	tga
100408	3AORF030	2	6299..6694	131	tataaaaaagggaggttggcagataa	atg	tga
100409	3AORF031	1	20833..21225	130	tttaaaaaattataggggtgagaaa	ata	tga
100410	3AORF032	-2	39984..40361	125	aaatagctgttagaggtgtaccctt	ata	tga
100411	3AORF033	1	7957..8325	122	gaatatctgcgtctttttttatttga	ata	tga
100412	3AORF034	-2	28506..28871	121	gttatcaacttaaggaggtgataac	atg	tga
100413	3AORF035	-2	10671..11036	121	tcttagcttcttaaacagcaccgcca	ata	tga
100414	3AORF036	2	30020..30382	120	accatatttttaaggaggagtttaatca	atg	tga
100415	3AORF037	2	21818..22165	115	aagtgtaagtaaatagtttaagagtc	gtg	tga
100416	3AORF038	-2	42603..42347	114	gtactcaactttcaactgcttcaacc	atc	tga
100417	3AORF039	2	21386..21727	113	tccagaaaaatttagagtcataaggtt	ata	tga
100418	3AORF040	-3	29654..29995	113	ttgataaactcctccttaaaattgg	ttg	tga
100419	3AORF041	-1	4333..4671	112	tactaaattctacatctgacctatga	att	tga
100420	3AORF042	3	5568..5900	110	taaaaaagtggttaggtgatttttaa	atg	tga
100421	3AORF043	1	25690..26019	109	tacaaatttaatatagctcttcgcat	ata	tga
100422	3AORF044	3	29676..30005	109	gtcttaataattatataaaggagtt	att	tga
100423	3AORF045	3	30..353	107	cgctagcaacgcgggataaatttttc	atg	tga
100424	3AORF046	3	27894..28214	106	aagatattgaaaagctaatctcccc	ata	tga
100425	3AORF047	-2	11907..12227	106	ttcgccgcgaataatgattagcattt	ctg	tga
100426	3AORF048	-3	40343..40663	106	ccataacacatacactgtatgatct	ctg	tga
100427	3AORF049	-3	6749..7069	106	tgttaaacacattcttcagattctcca	ata	tga
100428	3AORF050	1	42700..43014	104	ttatgcaatcaaaaggtgttaagag	atg	tga
100429	3AORF051	-2	13077..13388	103	ttgtacgttaactccacacatcgccg	att	tga
100430	3AORF052	-3	3722..4024	100	gcatttctattctccttaataaactc	att	tga
100431	3AORF053	3	17145..17444	89	tggagacaatggatataaggagtggt	att	tga
100432	3AORF054	-1	19915..20211	88	ataatttatagcttgcgaatacataa	ata	tga
100433	3AORF055	-1	42436..42729	97	aaatcgtattgatattgattacagcc	atg	tga
100434	3AORF056	3	40455..40745	96	taaaattttgtatacaagggtgaataa	atg	tga
100435	3AORF057	-1	38665..38952	95	atcatcacgcgtcttgcattgacgt	att	tga
100436	3AORF058	-1	21265..21549	94	gaattttctatcttaacttgcataa	att	tga
100437	3AORF059	-2	10278..10562	94	cttagccgcgttcccaactgcacgt	att	tga
100438	3AORF060	1	5278..5556	92	atatcagccgaatcaggggtgatgaa	atg	tga
100439	3AORF061	1	35668..35946	92	tttgaagaagagaggttgattcaa	ata	tga
100440	3AORF062	2	35912..36187	91	gttaaatctggaaatggaattaaaca	ata	tga

100441	JAORF063	3	36720..36995	91	cggaagtagcggagtgtaaaagacat	att	tga
100442	JAORF064	-2	35694..35969	91	ccgttatagcgctagcactaataa	ctg	taa
100443	JAORF065	-2	32697..32972	91	aacogtcttctctttgttaaatagg	ata	taa
100444	JAORF066	3	29157..29429	90	caaaccttaacatttatctaaagga	gtg	tag
100445	JAORF067	-2	26661..26930	89	atactcttttagcggaatcggtatga	ctg	taa
100446	JAORF068	-2	9624..9893	89	ttttaatgcactctcccatgtattga	ata	tga
100447	JAORF069	-3	13847..14110	87	tgcatttctctctgattcgtgttga	atc	tga
100448	JAORF070	1	34993..35250	85	tttagctccaagggcttttgactt	gtg	taa
100449	JAORF071	2	34745..35002	85	aaatgttcaagaatggagtgaagc	ata	tga
100450	JAORF072	-1	27379..27636	85	tttgctgttctctctttaaagtgtt	ttg	taa
100451	JAORF073	2	37367..37615	82	tggtaatagctattatcatttttga	att	taa
100452	JAORF074	-2	23466..23714	82	cgcttctctttttaaattcaatat	att	taa
100453	JAORF075	-3	2471..2719	82	agctactgtttaaattcttcaaac	ttg	tga
100454	JAORF076	1	26047..26292	81	aagtagcttttcttggcggggaggt	gtg	tag
100455	JAORF077	2	28292..28537	81	aacatcttaaaaggaggaataacaa	atg	tag
100456	JAORF078	-1	5836..6075	79	ttttgtataaggcttagattttagtc	att	taa
100457	JAORF079	-2	5460..5699	79	attcagtcgctttttaaattctctc	atc	taa
100458	JAORF080	-2	31350..31586	78	cctgttaactctttagttttttaa	ata	taa
100459	JAORF081	-3	8252..8488	78	aagtcttcttaaatccgtacctga	atg	tga
100460	JAORF082	-1	35905..36138	77	atatttatagacaacttgaccgtc	ata	taa
100461	JAORF083	-1	34039..34272	77	atagttcaactggattattaaata	ata	tga
100462	JAORF084	-1	12007..12240	77	acatttttttcttctcgcgcacaa	atg	taa
100463	JAORF085	-2	32367..32597	76	cttacaaggtatagagaataacga	att	taa
100464	JAORF086	-2	30618..30848	76	atataacttaagttgaggtattct	ata	taa
100465	JAORF087	-3	24746..24973	75	ataggcttttaagttcaacctcttca	atg	tga
100466	JAORF088	-3	12980..13204	74	tccttcttttctcgtaccaccatgga	att	tag
100467	JAORF089	3	4290..4508	72	acaggagaagcttatcaatctttaa	atg	taa
100468	JAORF090	3	28926..29141	71	ttatacacgaaggagcataaacaa	atg	taa
100469	JAORF091	-2	13587..13802	71	cttgccttctgaattgcttagataa	atg	tag
100470	JAORF092	2	26471..26683	70	aaacgaacacaaaggagggttca	atg	caa
100471	JAORF093	-1	2524..2736	70	tcacacgttttctcatagtagtct	ttg	tga
100472	JAORF094	-3	25334..25546	70	tggtcttttaataataaaagcgtct	att	tga
100473	JAORF095	3	8316..8525	69	aagagatgggaagacagagaacaa	atc	tag
100474	JAORF096	2	36992..37198	68	aaacagttcaaggaggtcatgagga	atg	tga
100475	JAORF097	-1	32593..32799	68	aaagcttaataacctctgtctgtat	atg	taa
100476	JAORF098	-1	15346..15552	68	aatccatttaaatcacctacataaa	ata	tag
100477	JAORF099	1	7225..7428	67	actggtgactggatgaacagaaag	ttg	tag
100478	JAORF100	-2	22620..22823	67	ogacttcatgaccggcatgtcttaa	ata	taa
100479	JAORF101	-1	40060..40260	66	aaacctacagcgaggaagggaagag	gtg	taa
100480	JAORF102	-1	35035..35235	66	ttctatctctcttaaaataaagttag	ctg	taa
100481	JAORF103	-2	1149..1349	66	atttttttggaggtgtgggttaatca	ata	taa
100482	JAORF104	1	27661..27858	65	aaacaacttaaggaggaacagacaa	atg	tga
100483	JAORF105	-2	9420..9617	65	gccttaagtcaaacgctgtgattagac	atg	tga
100484	JAORF106	-2	23244..23438	64	caccagtaattcttgaattagttga	ata	taa
100485	JAORF107	2	11966..12157	63	ctctaaaaagatgctgttagtagacg	ttg	taa
100486	JAORF108	-3	35054..35245	63	ttttcatcatttctatctccttaaa	ata	tag
100487	JAORF109	-3	16010..16201	63	gttcttaattccaatgtactgacag	ttg	taa
100488	JAORF110	-1	6184..6372	62	attttcagtgactttatattagtagt	att	taa
100489	JAORF111	-2	16500..16688	62	gtagtcacaattgctttgtatttga	ttg	tga
100490	JAORF112	-2	8502..8690	62	cttaattctcgcctgatacttttcc	att	taa
100491	JAORF113	1	34162..34347	61	tatgaaggattaggaggtgtgattgc	atg	tga
100492	JAORF114	2	12356..12541	61	ggatatcacactaaggcttatagcta	ata	taa
100493	JAORF115	-2	7635..7820	61	tgaagtctccctcagctacacggtag	att	tga
100494	JAORF116	-1	26434..26613	59	tttagcttctggaagtgtgtaaatct	ctg	tga
100495	JAORF117	-3	17804..17983	59	atagccattatttctagctgtgtgc	atg	tga
100496	JAORF118	2	27899..28075	58	attgaaaagcttaatttcccataaag	att	caa
100497	JAORF119	-1	39268..39444	58	acgaataccggtcaacttgttttagat	atg	tga
100498	JAORF120	-2	37152..37328	58	tagctattaccatgaacttcagct	ctg	taa
100499	JAORF121	-2	18900..19076	58	aaggtagctctctccatttaccact	att	taa
100500	JAORF122	-1	21550..21723	57	taagcatggtaatacactccttttaa	atg	caa
100501	JAORF123	-3	33062..33235	57	aaacgttcttctttaaataagatctc	ttg	tag
100502	JAORF124	2	21212..21382	56	aaattagaagggttaaaggagaga	ctg	tag
100503	JAORF125	-1	22051..22221	56	aaatcaggattgaactgtctccta	atg	tga
100504	JAORF126	-2	7821..7991	56	tggttttctgtgttttaaggtcttta	att	tga
100505	JAORF127	-3	34712..34882	56	ttgcattacctattgcgaatgctag	ttg	taa
100506	JAORF128	-3	24056..24226	56	tttttaaaatcaagcgtcttctgtt	ata	-taa
100507	JAORF129	-3	4940..5110	56	cataccatgcagttatataaacaa	ata	tga
100508	JAORF130	3	27171..27338	55	cagaatttaactatcgatgatttctga	atg	caa
100509	JAORF131	-1	40387..40554	55	ccttctggcataataataattctat	ctg	caa
100510	JAORF132	-2	1860..2027	55	gcgataacattcactccttaacgc	att	tga
100511	JAORF133	-3	42317..42484	55	acaaagtcttcttctgtattgtagtaa	ctg	tag
100512	JAORF134	2	12671..12835	54	tcatacaaatctttaaagggttga	ctg	tag

100513	JAORP135	-1	39484..39448	54	ataatagctattagcttctgcccag	att	taa
100514	JAORP136	1	29710..29871	53	accttacaacaaaaatactatcac	att	taa
100515	JAORP137	1	37186..37347	53	ggcagctgtgttgaataataaagga	gtg	taa
100516	JAORP138	2	20996..21157	53	aatggggaaatagtttctaacgaag	att	taa
100517	JAORP139	3	15114..15275	53	tcaactgaattgaagtaagtttaa	atg	taa
100518	JAORP140	3	29442..29603	53	aaaatgggtattaggaggattatcaa	atg	taa
100519	JAORP141	-1	39883..40044	53	cacaccataatcttttccaaatcga	att	taa
100520	JAORP142	-1	20416..20577	53	accacctggaaggtcccatataaa	att	tga
100521	JAORP143	-1	1942..2103	53	ataaagcttagaagctgactgaca	atc	taa
100522	JAORP144	-3	39380..39541	53	ttccaccagtttcaatctcttaagaa	atc	taa
100523	JAORP145	3	20388..20546	52	cttgagtggtcagaattagctatta	atg	taa
100524	JAORP146	-2	2358..2516	52	aacatgtccatattatgaacaatca	att	tga
100525	JAORP147	-3	5606..5764	52	gtgatttgtttgtgttagatattca	att	tga
100526	JAORP148	2	34145..34300	51	tttacttctcgtcttatatgaagg	att	taa
100527	JAORP149	-1	7918..8073	51	tatttctcttgatttactaatcttaa	ata	taa
100528	JAORP150	-2	11745..11900	51	ttcatccttatgtcttctgactcaga	ata	taa
100529	JAORP151	-3	7097..7252	51	tttaccctcatgatacccggtataca	ata	tga
100530	JAORP152	1	21652..21804	50	ctaaaaatattagagacatcaagat	gtg	taa
100531	JAORP153	2	5381..5533	50	tcggctaaagctgaattactattaa	gtg	tga
100532	JAORP154	-1	39670..39822	50	ttgataaaaatcgtcttcttcaagg	ata	taa
100533	JAORP155	-1	38233..38385	50	ataggctctacaaaatgcacaca	att	tag
100534	JAORP156	-1	33040..33192	50	tatctgaataataatgctttgcttaa	att	tag
100535	JAORP157	-2	10119..10271	50	cttcaatgatttgcctatagctatta	att	tga
100536	JAORP158	-3	36074..36226	50	atcogtcttatgatttcgttctgpc	att	taa
100537	JAORP159	-3	18338..18490	50	taaaatagtttcttatatttggatta	ctg	taa
100538	JAORP160	3	39399..39548	49	gtttggttaatggaaatggcagaac	ttg	taa
100539	JAORP161	-2	8976..9125	49	ttgtactctttagtttttgaacttga	ttg	tga
100540	JAORP162	-3	31199..31348	49	ttgttaactatcttcagggttataac	ctg	tga
100541	JAORP163	-3	14459..14608	49	attatccttgagaagaacagtttga	atc	tga
100542	JAORP164	3	25182..25328	48	tttttctttagcttttcttgacaaa	gtg	tag
100543	JAORP165	3	28353..28499	48	aatcttcttcttatgacaggaag	att	taa
100544	JAORP166	-1	8899..9045	48	gtactgcgtcacttgccttttttag	ttg	taa
100545	JAORP167	-2	411..557	48	taatacaagttgaogtttagatcct	ttg	tga
100546	JAORP168	-3	25973..26119	48	gctgagtacttcaatgtgaagatta	atg	tag
100547	JAORP169	-3	25151..25297	48	aaaaaaacgctcacaagtttagagc	ttg	tga
100548	JAORP170	-3	24995..25141	48	taagaaaaaagatttagtattcttc	att	tag
100549	JAORP171	1	23437..23580	47	aaaggttaataacgtaaggaagcgt	att	tag
100550	JAORP172	2	32414..32557	47	ctatttgacctgctgtataaaaggt	atg	taa
100551	JAORP173	-1	38005..38148	47	ataagttgtatcatcgaagtaatcc	atg	taa
100552	JAORP174	-1	4123..4266	47	atttaagatttgataagcttctctct	gtg	tga
100553	JAORP175	-1	3124..3267	47	ttcatttgaaaatacttagctttca	ttg	tag
100554	JAORP176	-1	580..723	47	cattttctccatcttctgatacagc	ata	taa
100555	JAORP177	-2	39819..39962	47	ttagaatatcttcttaatttccatag	atc	tag
100556	JAORP178	-2	38466..38609	47	tttagctcttcatcttgagcaccat	ata	tag
100557	JAORP179	-2	33927..34070	47	ttttgcccacacacaggtctatttc	atg	tga
100558	JAORP180	-2	33555..33698	47	cgcttttccgggttttacagttatta	att	tga
100559	JAORP181	-2	29538..29681	47	atagttatttttctgtgaaggtcat	att	tga
100560	JAORP182	-3	17099..17242	47	aatatcactactgctgcataaggt	atc	tag
100561	JAORP183	2	23750..23890	46	ttaaaaaaacaaagcttttagtat	ata	taa
100562	JAORP184	-1	31648..31788	46	tggaaatttccagatttgcaggaact	ttg	tga
100563	JAORP185	-1	30565..30705	46	attttgtttcaataaagctattac	atc	tag
100564	JAORP186	-1	16951..17091	46	gagaattccaagctactagtgataaa	atg	tga
100565	JAORP187	-1	7153..7293	46	tatccaacgaatacttttttgaa	att	taa
100566	JAORP188	-1	1237..1377	46	ccagctcttctaaagaacaatttc	att	taa
100567	JAORP189	-2	33309..33449	46	catttgagaagcggatgctcatat	atc	tga
100568	JAORP190	-2	7197..7337	46	gtacgaacttgcagactcctctga	atg	taa
100569	JAORP191	-3	41459..41599	46	tcatctgtataaactgcacgcttag	ata	tag
100570	JAORP192	3	4863..5000	45	gatgctattattaaagcttttagcag	att	tag
100571	JAORP193	3	25965..26102	45	tatacgaactagtttagactctt	ata	tga
100572	JAORP194	-1	37069..37205	45	ctagtgaataataaactcttagtat	ttg	tga
100573	JAORP195	-1	11749..11886	45	tttgatcagcaatagctataaattt	atc	tga
100574	JAORP196	-2	40764..40901	45	atctttagcaacttgttttaggtgct	atg	tga
100575	JAORP197	-2	31989..32126	45	ggttaaaacccccactattgactt	ata	tga
100576	JAORP198	-3	36431..36568	45	tttattttatgacataaactaccattc	ata	tga
100577	JAORP199	-3	33515..33652	45	ttccaaaaattaaactatgttaggat	ttg	tga
100578	JAORP200	-3	21233..21370	45	ataagattataacctatgaccttag	att	tga
100579	JAORP201	1	23293..23427	44	aaqccataggtgtgtgatatcta	gtg	taa
100580	JAORP202	-1	39088..39222	44	atagtcaaaatttacctctggtccc	att	taa
100581	JAORP203	-1	16309..16443	44	ttgtcttgccgcttaaaatcaactt	ata	tga
100582	JAORP204	1	23845..23976	43	atgtttattatcaatcaaatataa	att	taa
100583	JAORP205	1	29500..29631	43	gtgtgtgtgctcagggtcttagcga	ttg	tga
100584	JAORP206	2	16667..16798	43	gaaaaatcaacggtcttaaaattaa	ttg	tag

100585	3AORF207	-1	35386..35517	43	tgcagatttatagactccttcttga	atc	taa
100586	3AORF208	-1	30013..30144	43	cagttgagctgtttcattcttttggc	att	taa
100587	3AORF209	-1	28366..28497	43	taattcctgtgcttctcagttgggtt	ata	tga
100588	3AORF210	-1	15739..15870	43	catcaagcttatttgcactcga	gtg	tag
100589	3AORF211	-1	7693..7824	43	taactgaagttccctcagctacacc	gtg	tga
100590	3AORF212	-2	4314..4445	43	ggttctgaaacaatttctttagaaa	gtg	tag
100591	3AORF213	-2	4011..4142	43	tgtttgatgtcttccatataat	ttg	tga
100592	3AORF214	-2	1722..1853	43	tctgtctagtcttcaactgaacatta	ctg	tga
100593	3AORF215	-3	16616..16747	43	tcttcatttgtttgcgtattagcat	atc	tag
100594	3AORF216	-3	15833..15964	43	gtcattttgcacgaagtcttttctg	ttg	tga
100595	3AORF217	3	6363..6491	42	gatgcagagctcccaacataataa	att	tga
100596	3AORF218	-1	32146..32274	42	aataagctataattagatttoga	atc	tga
100597	3AORF219	-1	29800..29928	42	ctagggctcatcactttgttcttta	atc	tga
100598	3AORF220	-1	18409..18537	42	gcattaacctgatacgtcttcttcc	ctg	tag
100599	3AORF221	-1	13234..13362	42	ttttatcgcttaaccaagatgcacc	atc	tag
100600	3AORF222	-1	12313..12441	42	cccaagctttatctgagggcatctga	ata	tga
100601	3AORF223	-1	4915..5043	42	tccatcatagttaattccaaaataa	ttg	tga
100602	3AORF224	-1	2125..2253	42	attaactactttataatcttcatac	att	tga
100603	3AORF225	-2	26298..26426	42	tcgtttgtaacaacttgatttaaga	ata	tga
100604	3AORF226	-2	17184..17312	42	cgcttatcttttaatttatctaat	atc	tag
100605	3AORF227	-2	1425..1553	42	atcttcttccattctctatagggc	att	tga
100606	3AORF228	-3	31055..31183	42	cattttttgatgacagcagtttat	ata	tga
100607	3AORF229	-3	22592..22720	42	gtataaccatgacggctacaaagc	ata	tga
100608	3AORF230	-1	27883..28008	41	gaaggcagggctggttcttggatta	ttg	tag
100609	3AORF231	-2	29988..30113	41	gcttcttctaactttcttctgataa	ttg	tga
100610	3AORF232	-2	22465..22590	41	tatctgggaatttaattcaataaa	ata	tag
100611	3AORF233	-2	9264..9389	41	aagtttgcgaagtgtactttgagct	atc	tga
100612	3AORF234	-3	23033..23158	41	accttaactcagataagcgataatt	ata	tga
100613	3AORF235	1	25558..25680	40	aacactgctgaatagacgctctttt	ata	tag
100614	3AORF236	1	34420..34542	40	acattgagagaagtttcagaaanaa	atc	tga
100615	3AORF237	3	38442..38564	40	gaagaaagctataganaactttatcc	ctg	tga
100616	3AORF238	-1	33628..33750	40	caatcattagaaaacctttttctat	ata	tga
100617	3AORF239	-1	29248..29370	40	tcttcttaatttagaanaatttaata	atg	tag
100618	3AORF240	-2	18156..18278	40	gtcttctcaattctgtatagaatttt	att	tga
100619	3AORF241	-2	8088..8210	40	tttcaaggtcttttctataagtttta	gtg	tga
100620	3AORF242	-3	39149..39271	40	ttagcaaaagcagatttaactacacc	ttg	tga
100621	3AORF243	-3	23558..23680	40	aaaatttaactgtttataatttttaa	ata	tga
100622	3AORF244	-3	1697..1819	40	catttcttaaaaggtattattatga	ata	tga
100623	3AORF245	1	19015..19134	39	agttatgcaaggaatgatgactt	ttg	tag
100624	3AORF246	1	22504..22623	39	gctaatctaaacactttcacatcgt	ttg	tga
100625	3AORF247	-1	40567..40686	39	aaagtatttactgttcttttattcc	ata	tga
100626	3AORF248	-1	23956..24075	39	tttagatttctgaagaaagtagca	ata	tga
100627	3AORF249	-1	11113..11232	39	cacctttcccaaacacttttacagt	atc	tga
100628	3AORF250	-1	8719..8838	39	ttttatagcttctactagtctttaa	ata	tga
100629	3AORF251	-2	16899..17018	39	aactcgtctgttaagcgttgttga	att	tga
100630	3AORF252	-3	37025..37144	39	acaactgccttaatttaataactgc	att	tga
100631	3AORF253	-3	29138..29257	39	tctacatactccaaacaattgatgg	att	tga
100632	3AORF254	-3	15476..15595	39	caaatcaattcattaaaatccatta	ctg	tga
100633	3AORF255	1	13552..13668	38	ttatagacaaagtaaaaatcgtggt	ttg	tag
100634	3AORF256	2	12545..12661	38	aaaagtgcgaagggctggttaacgg	ata	tga
100635	3AORF257	2	41870..41986	38	gggcatggatttaacttacaacag	gtg	tga
100636	3AORF258	3	10827..10943	38	tcaaaacttttgaanaacggtttagg	att	tga
100637	3AORF259	-1	34570..34686	38	gtgacatcgaaacagtcaggtcac	gtg	tga
100638	3AORF260	-1	12389..12505	38	aagcaggttagcccaatcagcattga	att	tga
100639	3AORF261	-1	23830..23946	38	cctttttaacttttaacaaatttaa	ata	tga
100640	3AORF262	-1	8158..8274	38	ccatctcttctgttctcagtttctga	atc	tga
100641	3AORF263	-2	14001..14117	38	ttataccctgcttttccctcgtatc	gtg	tga
100642	3AORF264	-2	294..410	38	tttgcgtgttttatttcccttga	gtg	tga
100643	3AORF265	-3	42683..42799	38	tgacaagaataattatctctatcta	atg	tga
100644	3AORF266	-3	31979..32095	38	aatccctcatcagtgcttaatttc	atc	tga
100645	3AORF267	-3	26306..26422	38	ctgtaacaacttgatttaagaatac	atc	tga
100646	3AORF268	-3	16490..16606	38	tacatacaaggttagcttttttat	ttg	tag
100647	3AORF269	-3	9872..9988	38	tgaagaccttctaacctgagttag	ata	tag
100648	3AORF270	1	21829..21942	37	atagttaaagtcagtgcttcggga	ctg	tag
100649	3AORF271	2	29468..29581	37	tgaagcacaacataaaaagctacct	att	tga
100650	3AORF272	3	2955..3068	37	gagttaaacagattttacttgcagc	ata	tga
100651	3AORF273	3	5010..5123	37	tttggcaaacacagtagtatttacag	atg	tga
100652	3AORF274	3	19956..20069	37	tcaagtatagatgaattaaagcaac	ttg	tga
100653	3AORF275	3	39882..39995	37	gatattgttaccacaggaatgtag	att	tga
100654	3AORF276	-1	27211..27324	37	attaaagtgcgttatttaatttagat	att	tga
100655	3AORF277	-1	13516..13629	37	cgacctctattaagtttaagtcac	ctg	tga
100656	3AORF278	-1	11893..12006	37	ttttatatacacgaccactggataa	atc	tga

100657	JAORF279	-2	17535..17640	37	tttctaaagatttcttactgctgc	ttg	tta
100658	JAORF280	-2	6474..6587	37	tcaaaataagcatcttaactgactag	atg	tta
100659	JAORF281	-2	759..872	37	ttttgatactggttgcgtcataatgg	att	tga
100660	JAORF282	-3	36608..36721	37	cccataaactccttgcactgactcta	ata	tga
100661	JAORF283	-3	14960..15073	37	tttcagttgaagaaacactcttttaa	att	tta
100662	JAORF284	1	18959..18969	36	atgttaacagagccaggtctttact	att	tta
100663	JAORF285	2	8237..8347	36	aaaacttatcaaaagccttgagag	ata	tta
100664	JAORF286	3	5157..5267	36	tatgatcagcaactgacattagaca	gtg	tag
100665	JAORF287	3	38610..38720	36	tttgatttagtagcgcatacacttat	atg	tta
100666	JAORF288	-1	36454..36564	36	tttatgacataaactaccattcctac	ata	tga
100667	JAORF289	-1	30217..30327	36	aacaattttttctaatgctcttct	ttg	tta
100668	JAORF290	-1	16678..16788	36	gcttttttgcataattctaacagctt	atc	tga
100669	JAORF291	-2	14310..14420	36	gtctagtttaaggggataaccatctc	ctg	tga
100670	JAORF292	-2	11457..11567	36	ttctttcaactcttctgattttctga	ctg	tga
100671	JAORF293	-3	29462..29572	36	ttcatanaagttattccttatanaat	atg	tga
100672	JAORF294	-3	22388..22498	36	accatttccattttggccaaacgat	gtg	tag
100673	JAORF295	-3	18629..18739	36	aaaaggaaagccttcttgagtgaagt	att	tag
100674	JAORF296	-3	6332..6442	36	catcagacatgaggtctgaaggtaa	atc	tta
100675	JAORF297	1	13984..14091	35	aaatggttgaagtcacttaaaagta	gtg	tag
100676	JAORF298	1	40174..40281	35	tatcaaatgttgcactcttttttga	gtg	tta
100677	JAORF299	2	1481..1588	35	gcggcggtgctgacttttgcgttag	ata	tta
100678	JAORF300	2	40451..40558	35	aataataaattttgtatcaaaagtga	ata	tga
100679	JAORF301	3	25479..25586	35	accactagttcaaaacttctataact	ata	tta
100680	JAORF302	3	32106..32213	35	gaagatgatttctgagtgaattagaca	ctg	tga
100681	JAORF303	3	36024..36131	35	gacacagaggggatttataaaagaga	ttg	tag
100682	JAORF304	-1	37762..37869	35	accgacaaactccgcaacactctttt	ata	tga
100683	JAORF305	-1	24088..24195	35	ttctcttttaacaaactcaaaactga	ata	tga
100684	JAORF306	-1	19507..19614	35	atcattaggttaactgaacttttaa	ata	tga
100685	JAORF307	-1	16081..16188	35	atgtactgacagttgcagatacagt	atc	tag
100686	JAORF308	-1	11398..11505	35	ttctcttagttctgattanaattgtt	ttg	tta
100687	JAORF309	-2	33003..33110	35	aaacagaccccttaccgcttcatca	ctg	tta
100688	JAORF310	-2	24894..25001	35	gtaaatcgaaactcgtaccagctga	att	tta
100689	JAORF311	-2	22005..22112	35	ttcgtagtggtcattacttctttaa	ttg	tag
100690	JAORF312	-2	21711..21818	35	aaaaataaaagccagtgccgaagca	ctg	tag
100691	JAORF313	-2	17901..18008	35	cattaggtcttagacgacttagcat	ata	tta
100692	JAORF314	-2	16710..16817	35	taattcagttcttagaggtatcattt	att	tag
100693	JAORF315	-2	15990..16097	35	acatatctctgatacttctggtaa	att	tag
100694	JAORF316	-2	2862..2969	35	aattcttctctatactgttttagaga	ttg	tag
100695	JAORF317	-3	40217..40324	35	tccttaaacactactttttaaacttt	ata	tga
100696	JAORF318	-3	37535..37642	35	tggtgggctctctcttcttattttaa	ata	tta
100697	JAORF319	-3	34421..34528	35	ttcttctatcttttatttgcactctg	ata	tga
100698	JAORF320	-3	28262..28369	35	catttctgttgatattctttagttcg	atg	tga
100699	JAORF321	1	23989..24093	34	taaaaagggttttaataaaaatgta	ata	tga
100700	JAORF322	1	34660..34764	34	aaagagaagattgagaccatggcttt	atg	tta
100701	JAORF323	3	30105..30209	34	ctaaataactgaactatcaactgtag	att	tta
100702	JAORF324	3	30258..30362	34	ggaaaaagagttcccttaaaaagcag	ata	tga
100703	JAORF325	3	40236..40340	34	gttgatcatcttttgggtatgcaac	att	tag
100704	JAORF326	-1	36964..37068	34	cgcatacaactgttaacctttga	ttg	tga
100705	JAORF327	-1	35242..35346	34	atttttgcctggtgataaattttt	ctg	tta
100706	JAORF328	-1	21916..22020	34	ccatttacctcttctgagatgttga	ttg	tga
100707	JAORF329	-1	18820..18924	34	ggtgggttaacttccaaagacacac	ctg	tga
100708	JAORF330	-1	15631..15735	34	ttatgaagtttttcaaaattagtaa	atc	tag
100709	JAORF331	-2	37998..38102	34	ttacgccaatagcttcatatctat	ctg	tag
100710	JAORF332	-2	7359..7463	34	tttataaacctttaaagtttttagtc	ata	tta
100711	JAORF333	-3	24584..24688	34	aaaaattataaaactataaaacct	atc	tta
100712	JAORF334	-3	24269..24373	34	tatttttaggtagataaattattaa	atc	tga
100713	JAORF335	-3	14273..14377	34	cacttcaagcaagttgacttttga	atc	tga
100714	JAORF336	2	7559..7660	33	gtaactttatctaaatttagaagcgg	ata	tag
100715	JAORF337	2	13277..13378	33	aataataggttaaaaagcaggagaat	ttg	tag
100716	JAORF338	3	9501..9602	33	taggacgtacagatgacgtgggct	atc	tta
100717	JAORF339	3	27348..27449	33	atatctaatataaagcgactta	att	tga
100718	JAORF340	-1	37372..37473	33	ttctatggttttctatcttatgaga	atg	tta
100719	JAORF341	-1	33421..33522	33	aagctaatctggacacttttccctt	ttg	tta
100720	JAORF342	-1	29047..29148	33	tttggactctctatcactcctttag	ata	tta
100721	JAORF343	-1	7549..7650	33	atgatacgcttgagactagaattgg	att	tta
100722	JAORF344	-1	7297..7398	33	ctgctgaactgttgcagattttga	att	tga
100723	JAORF345	-2	23850..23951	33	ttaaacctttttaaacttttaataaa	atc	tta
100724	JAORF346	-2	20607..20708	33	aaagatgtacagactagatttagtta	atc	tta
100725	JAORF347	-2	14175..14276	33	atctgttcttaaaagacgctaataa	ctg	tta
100726	JAORF348	-2	6984..7085	33	cgtaactggttgactgcttaaac	atc	tag
100727	JAORF349	-2	6882..6983	33	tgaagcagacataaactgtatttag	atc	tta
100728	JAORF350	-3	40748..40849	33	aactgcatttcaactaatgctgtaa	gtg	tga

100729	JAORP351	-3	38345..38446	33	ggcttagtagaatgttttcgtataa	atc	taa
100730	JAORP352	-3	38081..38182	33	tagttgaaggccaatcacattaacct	atg	taa
100731	JAORP353	-3	35432..35533	33	tagcattctcatatgatgcagattt	aca	taa
100732	JAORP354	-3	34952..35053	33	ttatcctgatacagatatctcttag	atc	taa

5

190

Table 9

10

Bacteriophage 96, complete genome sequence

15

20

25

30

35

40

45

50

55

```
1   catagttata ggcttttcag ctatatacca agataagatt tatcccgccg tctccataaa aatatgcttg
71  gaaacccctga ttttaatgggg ttttaactca gcaagtgtca aatatgtgtc agaaaaataa ttttctgaca
141 cgttgacccat gctctctttt atgttcatca agtaagttag agtaggtgtc taagtgtata gatattatat
211 aatggccctaa tcttttgcta atatatccaa taggrtatcc tttagaaggt aggaagatg tatgcgtgtg
281 tcttaatgaa taagggttta ttgtagtatc atttagtctt atttgactct tagcatgggt aatagacttt
351 ttaacggcat tatgacctaa ttttaaacac ttattstctg taogttttgg taattttgat aatttagctt
421 taatatgttg tatatctctt ttgtgtacct ccacagrtct gtccgogtta actgtttttg ttccacgaag
491 atgtattgta cctctctttt cgttttagact gatagggcac atattaarta catcgctgta tcttgacca
561 gtgatagcta ggaatgaaat aaaaataata ctogattcgt ctctagattt aaagtattct atcaattgca
631 agtattgttc tatgggtgat aatttagagt gttcgtcttt tgattttttt gtaccacgaa tatctatttg
701 atagctaggg tctttcttta aatagccctc atatactgca tctctgaagc attgtgataa acaactgttt
771 aatttacgaa ccgtttcatt agtaogacct cgaccgaatt cgttcaaaaa cttttgatac tccgaacggt
841 tgaattgttt tattaaaaaa tcactccoga aatattcgtt aaataatttt aatgaacgtt gataccataa
911 gaattgttgt gaagcgacat gtttcttatt ttttgaatct aaccaatcat tgcaatattc ttcaaacctt
981 ttttttctat cttaatttgt tccatcatcc aatctcttaa gcagttgttg agcagcgttg gttgctctag
1051 ctttagtttt gaactcctgac ttctttttct tctctgattt gaagaacgga tgbttttacg ctgactgcca
1121 agatgttgtt gctttattct tcttttttgt aattgtaaat gaagccattt tacttttctt cctcaaaatt
1191 ggcataaaat aatgaaggta ggcagactac ccgaattttt attgttgaac aactatttgt tcaactcttg
1261 tcttctctac tcttcttcta aaactatcat atgattgatt aggggtgtgt aacgacattc ctggaccacc
1331 tccagcatgt tgggttttgt ccggattatt tctcattctt tcaagtgttc ttttagactt taatatctct
1401 tcttaactag gttcgttttg gtogcgtgtt tgtgctgtgt gtcattattt ggtagctgga agattctctt
1471 gtacctgttg ctttagatgt ttattgttgt gttgactgtt gtttaatttt gtgtgtttct cgtgttttad
1541 ttgattartg ttatcgtttt gattactatt tcttttttct gcttctgctt tatctttagt ttctttctt
1611 ttgtctttgt tctctttctt tgtttcgttt tcttctgttt cctctttctt atcgccgttg ttgctacogc
1681 atgcacctaa cactaacgca ctacttaata ataaaaata taactcttct atgttttaca ctctttatt
1751 tgcattttgt tttataaat ctatgatttc attgttttgt tctatgattt tgttttctat tttatagatg
1821 tcttctnaca tctctattaa gacgaatttt tgatttatca tttcgtaggt aaacatttga cctgtgtgtg
1891 taggttaga aaaaacacta ctgaacgctg ttgaanaagct atctataat tgaccaactt tattttttaa
1961 taacatactt taccgctctt cagacattgt atttagttcg cgtttattta aagtttttct tataattttg
2031 tattttgttt cctgatttct ttogatttct tctacttcaa aagggatatt gttatttaac ttttogaata
2101 tatcagcttt ttcagaactt gacatacgat caaatacttg tttttgactt ttatttaact tccctogaat
2171 ttttccggca gtccaaagct ctttaactgt taacttatca ttaggaaact gattctcttt tcatatgact
2241 ccttttctca tatttcttta tatttaaaaa ctctcaacgg ctcaaatgta atogaactct cgcacatggt
2311 agttccaata ccgttatatct tcttatattg ttctattgco tccaatatgt attcttctgt taattgtaga
2381 tactcagaca actcatacaa gttacgttag ccataattgt aagctttctc aatttccgtt aacgggaactg
2451 ctgagataaa gccgtgtcgt cttgogtaat ttlogaactt gctgattgtt aatttlogatt gatctaaaaat
2521 gttgccatag ctcaacttgt ggtgggcaag ttctctatat aacttctcta atttgttctt ttoggataag
2591 gaaggtctaa taaaattttc tcttcttcta taccacacat cgaactcctg aggtactctt tgtgtttctt
2661 tcaactcaac ttcacatttc ataaacattt cttogtattt tcccatgogc caaacccctt tgggtcttta
2731 tttctttcta tctctaaccc attgcataaa attttcgttt tcttctcatt cttcgggagt aaattcactt
2801 ttatttgcgt gacggyctat ngtttcttga tgaatacttc tttcttctgt aattctcgat tttagtcat
2871 taaagtaarc tgcataattt tggacttttg atattctagg atatttaagt tcttttaagc agttagagat
2941 tgttgattga cttacocoga ttgcttcaga caattctact tgaagtaagt tgttctcttt cataagttgt
3011 tcttaagttct ctgataaaat tttctagca ctcttatatt ccaataattt ctctcttagt attacttaat
3081 gtaatactaa ttaacataaa gtaatatcac ttttcaatca aaaaatttat tttttgaaa taatatcac
3151 tttaggttgt gacataattc ttaagtgtat agtatagttg taaatgtcaa cgggaggtga taccgaatgc
3221 cagaaaaatt taaagagctc tctgtaaagg tctggagaac taattcgaat atgacacacac aagatgtcgc
3291 tgaataatta ggcgttacta aaactctgtt aatagatagg gaanaagatg acgcagaatt aaaaagctta
3361 caattgtatg ctttagccaa attattcaac acgaagtttg attatataaa ggcataaaaa attcaactt
3431 aattctactt taagtgaata aggaaggaac ccttcttctt taggtaagga tgttgcgtga attttagggt
3501 ccagtaagga aatctgaagt ggaaggaaga ccttcttctt taggtaagga tgttgcgtga attttagggt
3571 atgcacgagc agataacgct atacgcaatc acggaatctt ggaattatac gtttaacttt tgcagcttct
3641 gtcaggtcaa aacagaaata tgaatcatat cccgttagga aatctcaacg cgggttaact tggaggttt
3711 aaacaaagta aaacagaaaa cattagagaa accgttagga aatctcaacg cgggttaact tggaggttt
3781 taccgagctt aagaaaaact ggtgtctacc aagtaactag tgaaccaaag tgaaccaaag gatttaattt
3851 tgaagctaca gaagaaacaa aacagaaat taaaacgtg aaagatgatg ttattgatgt gaagaaat
3921 caaaaaactg atcggggaga ctacattttc caaaacacac gtagcgaatt attcagggat attcaattcg aagtgaaaaa
3991 gactacatgc gataacaaac caaaacacac gtagcgaatt attcagggat attcaattcg aagtgaaaaa
4061 gatgactggt gogagttcaa gaacgaacgt aagacaaaaa catttcgagc atgtaattga aatgatttgt
4131 aattggttcc cgtcaacagc tactttatcc agaatcaagc aatttgaatt gaatttttaa aacgaatat
4201 aggaagaggt gaatatggaa tacatcggat atgcagagc aatgogttt gtaaaaaata gttgctttt
4271 aaagatgat ctagagaaaa aagtctactc gaacaaagag tttaaaaaag aatgcattga cagatttgg
4341 cgaggacaaa agcgttatat aaaaattgac aaagctattc aatttatcgt taccatttta atgattatg
4411 aatacgaatt ataggaggag ttatcaaatg agtaaaactt ataaaagcta cctagtgcga gtaactatg
4481 tcacagctctt agcgattgta cttatgcctg tctatactt cactacagcy tgggtcaatt caggattcgc
4551 aagtatcgca acuttctat actacaaaga atacttttat gaagaataaa aaactcgtta cttgcgtcaa
4621 caagtaacag tgacaaacat ttatcaaat atacaactta attaaatcaa aatatacaga ggtagtcaac
4691 tatggctgaa aatattaaaa ctgaacacaa ttattacact aaagatttct caggatagag aatgaagaa
```

5

191

10

15

20

25

30

35

40

45

50

55

4761 gataactttg tagcaaatca agaattgaca gtaacaatcu cattgaacga gtacagaaaa cttattgaaa
4831 taaaggctgt taaagataaa gaagaaagata cttacagagg taagtatttt gcggaagaaa gaaaacacga
4901 aaaaattgaaa aaagaaaata taaaactaaa aaacaaaatt tatgaattac aaacgaaga agataacgag
4971 gaggaagaa agacaaagga ggaagaaac gatgtattac aaatttggtg agataaaaaa caaaattata
5041 agcttttaacg ggtttgaatt taaagtgtct gtgatgaaga gacatgacgg tatcagtato caaatcaagg
5111 atatgaataa tgttccactt aaatcgtttc atgtcataga tttaaagoga ctatatattg cgaaggatgc
5181 aatgcgtgac gttataaaacg aatggattga aaataacaca gatgaacagg acaaaactat taacttagtc
5251 atgaattggt agyaggtatg aaagtgaat gatatacaag agagagaatt agaaaacttc gaacaaagacg
5321 accgattcaa agtaactgat ctagacagtg ctactcgggt ttttaagaaa ctggatgcaa tcaactaa
5391 aggaattgaa atcaacgatt tgcataataa agaattgaa cgcataaacg aatgaaaga taagaagta
5461 gaaaattac agagtggcaa agaatattta caaagccttg taattgaata ttacagaata caaaaagaa
5531 aagatagcaa attcaagttg aatacaactt acggaaggt gacagccaga aaaggttcaa aagtcattca
5601 agtagcaat gagcaagaag tcaataaaca acttgagcaa cgaagtttg acaactatgt aaagtaact
5671 aaaaaacta gccaatcaga cattaagaaa gatttcattg taactgaaaa oggcacattg attgacgcaa
5741 acggcgaggt ttttagaggtt gtagcattg tggagaacac aacgtcatac acggraaggg tggaggaata
5811 gatgactgaa aaaaactaat aagatgtcga tattttaacg caactaggtg taagaagact cagcaacaaa
5881 aatgcacaaa ogttttataa atttgagata tccgggaagc tggtagctgg taagaactag ttttaacaaa
5951 aagatacaaa taccattaga ctagataaaa atggaggaag aacaaaggtt acagaagatg gggcagttgt
6021 gcagattaa agtaataagc atttttagtc agtgattaaa atgtgcctta aaattattga acaactaaga
6091 gaaaacgaaa aacaaattga tctgttagtg attgaaacaa tccaaaagtt acgtgatac actatggacg
6161 actatctgga cggtaaatca aagaaacoga catttaagta tgggggcagc tgtgtacac gcatgttaag
6231 tattattcgt tatatttcta aattacaaga actattatca tttcatcttg ctataagcgg acacggggcg
6301 attaaacaag acaaaagatg tgaaggaagc actatcaatc caacactcac gatagagga gatgaacaaa
6371 taaaaaaagc agtcactcag caatctgacg tgttagcagc aatgacata gaagaacatg agcaagcgg
6441 cgaaaaaact tatcaatag tacttaacgc tgaacctca aaattattcg agcaaaagat aagacttca
6511 agcaacatca aaattacaac caaacgtttc attaatcaaa gtattaaaga tttgttcaaa gcaattagaa
6581 atggttaatto aaattaat aaagggagc tataaaatt atgaaaacta ctggtagaac acaactacat
6651 caagaaacta atcaagaggc attcatgaaa ggtggggact ttttagagc tgggaattt acagtaaaag
6721 ttgcaaatgt cgaatttaac gacagagaaa acagatact cagatttgtc tttgaaaaa acgaaggtta
6791 acatcaaaa caaaccaact tgcctccacc attccaacaa gattatcaag aaaaaaata tatcaggtta
6861 cttagtagat taggaattaa attgaactta ccagatttaa cttttgacac agatcaatta attacaacaa
6931 tgggaactat tgaacttaaa aataaattta acgaggaaca aggcagatgt tttgtgaagc tctcatatgt
7001 aaaggttgg aataaagacg atgaagtagt taataaacca gaacctaaaa ctgatgagat gaacaaaaa
7071 gaacagcaag caaatggtaa acagacactc atgagtcncc aatcaaaccc attcgttaat gotaatggtc
7141 caatagaana caatgatgat gatttacgtc tctaggagct ggttttaagt caatcatta caagatacca
7211 gaaagacaat gacggtactt attcctcgtc tgcactcgtt gttgaacttg aacaaagca cattgattta
7281 taagaaaaacg gatattcgtc aaagcagaaa gtagaggttc cggacaaata aaaaactatc atagaacaa
7351 gcaaaaaaat atttcgcaatg ttagagataa tagaacttca ctggggcgaa ccagtagaat caactgaaa
7421 attattacaa acagaatttg aaattatgaa aggttatgaa gaaatcagtc tgcgtgactg ttcaatgaaa
7491 gttgcgagag agttaataga actgattata tctgttatgt ttcatatca aataactatg agtgtagaaa
7561 cgaataagtt gtttaagcga gataaagcgt tattatattg ggttacaatc aacgcgaact gtgttaatg
7631 cggaaagcct cagcgagacc tggcaccatta tgaagcagtc ggcagaggtt tgaacagaaa caagatgaat
7701 cagtcagaca aacatgtgtt agcactgtgt agacaacatc ataagtaaca gacagcaatt ggtgttaagt
7771 cgtttgatga taataatcaa tgcactgact cgtggatcaa agttgatag aggtccta aaattgttaa
7841 agggaggaana aatgaataag ttactaatag atgactatcc gatacaagta ttacogaaat tagctgaatt
7911 aatagggtta aacgaagcaa tagtatgca acaaatatc tattggctaa acaactcaa acatanatc
7981 gatggcaaaa ctgggtttt taattcttat ccagaatggc aaaaacaatt tccattttg agcgagagaa
8051 ctataaaaag gacatttggg agtttagaaa acaaaattt attgcattga ggttaacta acaaggctgg
8121 atttgaccgt cnaaatgggt attcaatcaa ttatgaacaa ttaaaacaa tagtggcagc accatgggga
8191 caaaatggcc cgaagatgag gacaaattgg cagatgcaa gaggacaaa tgaccggacc aatacctag
8261 actacacaga gactaacaaa catagagaga cagacgagct ctcaagtc ttttaagtata ttagtccaa
8331 tttagaattt atcaaaaacc ctttaaaagc agaacagtta gaacacgaaa ttaaatcatt taagcaagat
8401 cagttcgaaa tagtaaaagt cgtacogac tctgcaaaag aaaaacaaa aggtctgaat tactactaa
8471 ctgtattaaa gaactggagt aaagaagcgg tttcagataa agaaagtgtc gaaaacaaat tgaacctgc
8541 taactctaaa aaagaacta ctgatgatg catagcaaaa atggaaaaag aattgagtga tgaactatgc
8611 cgaatggcaa aacaaagca ttagaataa ttaaaaagt tggtagctga taacactcg attttgataa
8681 accaaagtta aaactgtgga ttgatgtatt aagtcaaac ggggtatttc aaaccaactg aaaaactgta
8751 gatggatata tcaacagtaa caaccggcgc cgcctaaccc taccagcaat catgctgaag gcaactaaaa
8821 aagttatcat tgaagcggta gacaaagaaa cgcctacaca ccaatggaaa atgcagaatg acccogaata
8891 tgcagacaaa agaaaatag cgcctagata cttcatgaat aagttggcgc aatttggggc cgaatacga
8961 tgaattacgg tcaatttga attgaaagca caataatcgc tacgtactt aaacaacggc agtactaga
9031 aaagataaga gttaaagatt acatgtttac gaacgaanag tttaaaact ttttaacta tgaatggac
9101 gtcggaagaa tagatcatca agaatctat ttbaagcna ctaagataa agagttttta gatgcagata
9171 ctataactaa actttacaac tccgaattca ttggatagc attctttgaa cgttatcaac aagaattatt
9241 ggaagattat caaatcaaca aagcgaaga attggttaact gattcaaac taactaacg aaaaagaa
9311 tttaataact tgattgatga actcaaggat ttaaaacaa ttactaacg aaaaagaa gaaaccaa
9381 agtttgttga ggaattgtgc gatgaattat acagcgatag cctaagaag caaattaga cgggttataa
9451 gctcatggat tacaataag ggggatttga gccgtcgaa ttaattcga tgcagcgcg tccctcagtg
9521 ggttaagacg gttttgcatt aaacatgatg ctgaacatag caaaaatgg atacaaaaa tcttcttta
9591 gttctgaaac aactggcaca tcaatttga aacgtatgt atcaacaatt actggtattg agtcaacaaa
9661 gataaaagaa atcaggaact taacgcgga tgaattaaat aagttaaaga atgcgattga taaaatcatg
9731 aaattaggca tgaattttc tgaataaagt aatatcacac cgaagatgt gcgaagcga gcaatggcg
9801 attcagacag gcaacaagtt atttttatag attatcttca actgaggtat actgatgca aagctgatag
9871 acgtgtagca gtgaaaaga tatcaagtg cttaaagata atcgttaacg agacagggcg aatcatcgta
9941 ctactttcac aactgaatcg tgggtcogc tctagacgg ataaaaagc aatgctatog gacatgaag
10011 aatcaggcgg aatgaagca gatgcaggt tagcgatgct accttaccgt gatgattatt ataaactgta

5

192

10

15

20

25

30

35

40

45

50

55

10081 cgaagatgac agtatcactg gcaaatctat tgttgaatgt aacatagcca aaaacaaaga cggcgaaacc
10151 ggaataattg aatttgagta ctacaagaag actcagaggt ttttcacatg aatataatgc aattcaaaag
10221 cttattgaaa tcatatgtag aagagacaaa gcaaaagcgac cagattgtag caaatgtata tatcgagact
10291 ggttgggggg tcaatagatt gttggacaat aacgagttat cgcctttcga tgattacgac agagttgaaa
10361 agaaatcat gaatgaatc aactggaga aacacacat taaggagtg taaaaaatgc cgaagaaan
10431 atattactta taccgagaag atggcacgga agatattaa gtcacaaatg ataaagacaa cgtaaatgaa
10501 gtttattcgc tcacaggagc ccttttcagc gacgaaaga aaattatgac tgatagtgac ctaaaacgat
10571 ttaaaaggcgc tcacggggtt ctatatgagc aagagctagg attgcgaagc acgataattg atatttagag
10641 gtggccaat gagtaaatc aatgctaaag aagttgagta caaaggaatt gtatttgata gcaagtaga
10711 gtgcgaatc taccatatt tagaaagtta tatgaatgac actaactatg atcgtatcga aatacaacog
10781 aaatttgaat tacaacctaa attcgggaaa caaagacgga ttactatatt agccgatttc tctttgtgga
10851 aggaaggaaa actggttgaa gttatagacg ttaaaagtta ggcgactgaa gttgccaaac tcaaaagcaa
10921 gatattcaga tatcagtata gagatgtgaa ttaaaagtgt atattgtaaa cgcctaataa cacaggtcaa
10991 gaaaggatgg tatatgagga cttagtgaaa gtcagacgta aaagaaaaag agaaatgaag tgattctaatg
11061 caacacaaag catatataaa cgcacaatt gatataagaa tacctacaga agttgaatat cagcattacg
11131 atgatgtgga taagaaaaaa gatcgcgtgg caaagcgctt agatgacaat cgggacgaat tactaaagta
11201 tgacaaata acataaagac atgcatatat agaggtgaaa taatgaagt tgaaaggaat attcgaact
11271 aaatttgaagg taactatggc tagagataac gtaagtgtcc aagatttgca caatgaact ggcgtatcaa
11341 gatcaactat tagtgatatt aaaaacgaaa aagctgagat ggttaactta aatgtattag ataaattggc
11411 agatgctcta ggtgttaagt taagtgaact attactaga aatcacaaac cgcacaatt agaggttgg
11481 attaaaaaag taattgtata gagggtgaat aatgagatc cgtaaagatt aacggttaac catataatt
11551 taccgaact gaataatgaa tgataaaaaa gaacggttta actcaggaa tgggtgcaaa aagagtaga
11621 ggtggtggg cytggttaga agccttacct gcacctatg gtatgcgtt agctgagat aaagaaattg
11691 tgattccaaa aatcagggag cyagagagca aagagcgtga aatggttagg caacgcgta aagaggtgga
11761 actacgttaag aagaagccac atttgcttaa tgcctctcaa aaacattctc gtcgtcgtga ctggttcgat
11831 gtcacttate acaaatgct caaagaatgg agtgaagcat aatgagcata atcagtaaca gaagaataga
11901 tttgaacaaa acgcaagaca atgttaaaac acgggagcat tacacatacg gcaacattga aattatagat
11971 tttatcgaa aggttcacgg acgatataca cctcaactag catcgcgaat aggttaatgca atcaataact
12041 tctctagagc accgttaaaag aatggtcatg aggatttagc aaagggcgaag tttacgtccc aaagagcttt
12111 tgacttgggt gagggttaac gatggcaacg caaaaacaaag ttgattacgt aatgtcatta cgggaacat
12181 tggagttaga agactgtgaa aaatatcacg acgaacaaag taagactatg agtcataaag aagttagcaa
12251 tgtattgaa aactataaga caagcatatg ggtgaaagag ctatataacg aatgcattgc gtttggctcg
12321 cctaattgtt aaagagagtg atgacatga acgatagcgc acgcaaaaga tacttaaac aatttttcag
12391 cttcaagaga tatctgtatc aagacaacga gcagtggaac catatccatg tagtaaatgg caattattac
12461 tttcacggac attataaac gatgtttaa ggcgtgaaaa agacatttga tactgctgaa gagctgaaa
12531 tatataaaa gcaacatgat ttggaatatg aggaacagaa gcaacaaact ttattttaga ggagatgaaa
12601 actaatggca agattaaaag aaaaagaag atgacgctac tcgaactggt ggaattgggca tggaaacatc
12671 ctgaacaagt tgaaagttaa gtgtttcaat cagatagaat ggcgcagctt gggaatgta gogaagtaca
12741 tttttcaact gatgggcatg ggttttatc aaaaagtagta acagataaag atatttttac tgtagaatc
12811 acagaggaaag tcaactgaaga tactgagttt gattgtctag tagaactaaa cgtattgaa ggttttgaan
12881 tatatgaaa tgattcaatc agagagtgta tagacggtac ttccagagcg ttttatatac taacagaga
12951 taanaactatg acattaaatt ggaagaatgg ggaagtgtga gtatgatgca aacctataaa gtatgtcttt
13021 gttcaagtt ctttgcatct aaatgtgatt ataaattaaa gaacatttat ttcgtgaaa gtaogaatga
13091 ggaanaagcc uogaacatgg tattaaact gattcgtaaa aagctccctg tcgaactgc aagatagaa
13161 gtcgaanaag tggaggcaat ataagtatc aaccaacaa agaaagatta attaatcca tgaanaaaca
13231 tggagctgaa aatgttgact ctatcactga tagagcaaat gcaatagac actttagagc tcaatcaaaa
13301 gtttttaaa agacagctga tagatcaag aagcaacgag atgagcttat cgaggatata gctaagttaa
13371 gaaaacgtaa cgaagagctg gagaacatgt ggcgcacagt caaaaatgaa ttgcttgaaa gatacgaaca
13441 tttactgttt aaaaatagag aactacaccc tgagagcaaa gcgaacagga taggagctct ctatatagga
13511 ggtcaaaaga ctgcagatat tatactgtcg cgaatggaag aactagagcg acaaatgag ttctcgaat
13581 ttttagggca aatggaggca gacacaaatg aataaccgtg aacaaataga acaatcagtg atcagtacta
13651 gtgcgtataa cgttaatgac acagaggggt tactaaaga gattgagac gtgtataaga aagcgcaagc
13721 gtttgatgaa atacttgagg gaatgacaaa tgcatttcaa cattcagtt aagaaggtat tgaacttgat
13791 gaagcagtag ggtttatggc aggtcaagtt gtttataat atgaggagga gcaggaaat gagatttagt
13861 gtaggagata aagtataaa ccatgaaaca aacgaaagtc tagagattgt gcaattggctc ggagattata
13931 gagatacaca ttataaactg cttgatgatt cagttattag cattatagat ttattacta aaccaattta
14001 tctaattaa ggggacgagt gagtgaatg gaaacgatta aaaaatgtgg tgcgcaccc agttatcaa
14071 aataaaaatt taagtccgt ataatgaaa aagataatg tgaagaggt tcaaaaagaa ttaggtttct
14141 ttgaaatttt taatgaagaa gtgtattaa ctggattttt atcatttcaa aggtaccta ttacattat
14211 ttgattaat cctaaactc ataaagcgc tagatattac tttgtaacg agcatgagat tgaagatat
14281 ttgaatttt tggaggagca gtaaatgctt gaaatcatc accaacgtga tgcattgcta gaagaaagt
14351 atttaacga cgaactggtg tagagctag atattgtgt gaataaacgc aagtcagaaa atgaacagat
14421 tgatattgat agagtgctta aattattga ggaattaaa cgaataggga taacgaataa atgaataatt
14491 taacagtaga tcaattaaa gaactttac aatatacaaa ggaactcagc gatagatac cgaatagaaa
14561 ttttaattgac acagttagct tagatttat tgaatttgcg gagtgggtta acacactgca gttttttaa
14631 aattggaaga acaacacagg taagccatta gatacaaat tagatgagat tgcgtattc ttactttca
14701 gttgcaatt aactcagact attgttgatg aagaagattt ggaagagact actgaggtta tgggtgatt
14771 gattgaaat gaagtactt tacttaact acattcagtt tattttgttc atgtaatgca tacttaaca
14841 gaacaatttg taaaaggtat tgataatagt attgtacaa ttttaataat gcttttttg taoccaata
14911 ctactatad aatcgacaa ctactagcg ctatcaaaa gaaaatgaa aggaacacg aagacaaag
14981 tggacagca gaocgggaa aggtacagc gtaaaagac cttagatcga gtaagaggg ttttgggaa
15051 gtgcgcaat acttagcac aacttcaaa gattcaaac gacaaccaa tgaacattt actgctgta
15121 ggaataaca gaogtttaca gttgtgagg cggagagtag aagagagcg aagagagag acygaaca
15191 agttaagata aggaagatg gagtgcga agaaaacggt aacgattgat gtatgaaa acttattag
15261 agtgctagt aatgaatat cagaactatt atatgaatc gacagtgagt taatgtcago tgatgaagat
15331 ggcgataata gagatcga aaaaaaaga gacgattcaa acaagcctat acaaatatc gataaataa

5

193

10

15

20

25

30

35

40

45

50

55

15401 catgtcggagg aggcagacga tgaataacat acctaaaatg aaattcccca aaaagtacac tgaataatc
15471 aagaataata aaataaaac acctgaagaa aaagctaaga ttgaagatga ttctcctaaa gaaattaatg
15541 aaaaagacag tgaattttac agtccatga tggctaataat gaagtaacat gaattaaagg ctatgttaag
15611 aatgatgctt agtttaattg atactggaga tggcaatgat gattaaaaaa cttaaaaaa tggattgggt
15691 cgtatctctt attgctggaa tactgogatt attcggcgta atcgactga tggctgtgt catatcgctt
15751 atctatcacg tggctagtta ccaaaacaaa gaagratatc aagggaacat tacagataaa tataacaa
15821 gacaaataaa aagaagacag ttcttatatt tggtagacaa caagcaagtc atcgaaaact ctgacttact
15891 attcaaaaaa aaatttgata gogcagacac acaagctagg ttaaaagtag gcgcaaaagt agaagttaa
15961 acgattgggt atagaataca ctttttaatt ttatatccgg ttctatcaga agtaagagag gtatgaaat
16031 aatgattaaa caatattaaa gactattatt ctactagcg atgtatggc taggttaagta tgaactgag
16101 aaagtatata ttatgacgac ggctaattgat gatgtagagg cgcgagatga ctccgcaag ttgagcgatc
16171 agtctgattt gatgagggcg gagggtgtcag agtagatgta tagcaaaag tcaattgtta atgtatagg
16241 caacataaaa atgaagtgtat atgtattagc tgatgtaata ccggaatag atagcaattc atgtgcacag
16311 tatggcatac aagcaacgtt gccgaancca caaggggaaa actcaagtaa agttgaagat gttgtgtga
16381 ggcttgagag agcaaataaa aggtatgctc agatgttaaa agaggttgag ttataaaac atcgcaaca
16451 gagattggga cacygtgact ttgtctctt agagttattg aagaaaggtt atcaagggga tgcattatc
16521 aagaagatgc caactctaa attaaataga aacacttct tagcgcgccg tgaatgagta gcgaalaga
16591 ttatctact acagtgcaga aaatgacaaa aatgacagaa atgacgaaha tgacactatt tttaactgt
16661 gaattaaatt tacaataatt atttgaaga attatcttaa gacgtgggtt aatagccaca ttatgttct
16731 tcatcgatgt gattgagaag tgacaaaact ataaagatg atattgttac ctattaatca cctactacct
16801 gcttatatgg tgggtagttt aattcttgca tttagtgta taactatttt cctccttcca catttattg
16871 acgtagctcc tgcaacaagt gtgggggcat tttttatatt taataacta gaggtaataa cgtaaaggcg
16941 tgtatcacg tgaaaacaa ttgattaaat aacacggaag caagaaaagt ttgtgctagg actcatagag
17011 ggcaagagcc aacggyaaag atatatgac gcaaggtatt cgaataaagg taagatgggg gaattcttag
17081 ataaagagc gaggtaacct tttaaaaatc ggaaggttcc cggaaaggtac gaaaattgc gtcaagaagt
17151 agtgaacaa tcaaaatgga cncgccaana ggctttgaa gaataatgag ggctaaagaa tgaactaag
17221 aatgacattg aaatagaggg agtgaagaaa gcgacagctg atgcattcct cgtagttaa gatggtatga
17291 atagaaatgc gttaggtaac gaagttttag ctaaaaagaa aatagaact gaattaaaga tggctgagaa
17361 gaagattgaa caaatagata aaggtgacag tggaaacagaa gataaantca aacacttca cgaacataa
17431 acggaagtga tgcgcaatga ataaacttaa atcttlatat acggcaaaac aaattgaat attgaagcaa
17501 acgcataaac aagattgggt tatgttaatt aatcagggag caaagcgtac aggttaaaac atattaaaca
17571 atgacttatt ttacgtgag ttaattgcgt tgcaaaagt agcagacgaa gaaggaattg agacacctca
17641 atataactt gctgggtgaa cattaggtac gattcaaaaa aacgtactaa tagagttaac taacaaat
17711 ggacttgagt taattttga taactataat tcattcatgt tatttgggt tcaagtgggt cagacaggtc
17781 acgtaaagt aagtggtata ggaactatac gtggtatgac atcggttggg gcatatata atgaagcgtc
17851 gttagcgcat gaagaggtgt ttgacgagat taagtcaagt tgaatggaa ctgggtgcaa aatattggt
17921 gataccaacc ctgacctacc cgaactatgg ttgtgaaag attatattga aaatacagat cctaagcag
17991 gtatactgag taccacattt aagctcgatg acaataactt tcttaatgat agatataaag agtctattaa
18061 ggctcaaca ccatcaggtt tgcctatga acgttaactc aacggtatgt gggtgtctgg tgacgggtga
18131 gtatatgccc actttgattt gaatgagaat acgtataaag cagatgaact ggaacgata cctatcaag
18201 aatacttgc tgggtgcac tgggggttacg agcactatgg atctattgt ttaataggag gaggtataga
18271 tggtaacttt tatttattg aggagcagc acaccaattt aagttattg atgattgggt ggttattgca
18341 aaagattatg taagtatga tggcaattt aatttttact gogatactgc acgactgaa tacatcactg
18411 aatttgaag acatgatta cgtgcaata acgctgataa aagtaacta tgggtgtgg aggaagtgtc
18481 taagtgttc aaacaaaca agttactgt tcttatgat aatattgata gggttaagca agaggtattt
18551 aaattgttt ggacccctac aacgggagag cctataaag aatttgatga cgtgttgag tcgttaagat
18621 atgacctata caccataact attgctaag catattgagg ctctaataga gtacataaaa gacgtatag
18691 gatattgag cacaaggaat attgctaag catattgagg ctctaataga gtacataaaa gacgtatag
18761 agaaatggt taactctat aatagataga agacataat tgaactatga ccaatattca aacgtcgacc
18831 aattgaagaa aagaagatt ttgaactcgg tggaaatga agcgagtag acgtgtctgt taataacaaa
18901 cctacaact ctttgacag cgaattgtt gatcacggt tgggtattt acatgggtct cctgtactt
18971 atgatttga tgaaaacgca gaaaaaacg aaagtgtga aaagtttata accaacttg ccaatagaaa
19041 tagtggtag gatgaggatt ctgaaatagg taagtggca gcaatttgg gataggtgtc taggttagca
19111 tatattgata cgaatgggtga tattaggatt aagatataag atccctataa tgtattttt ttggcgaca
19181 atattttga acctacatac taattgogct actttatga aaagatgat gataatggca ctgattatgt
19251 gtacgogag tttacgata atgcttatta ttatgtatt cgaggagaag gtattgacg ttgcaagaa
19321 gttggagcat atgaacattt atttgattac aatccattgt ttggtgtacc taacaaacaa gagatgatg
19391 gagatgctga aaaggttatt cacttaattg acgcatatga tttaacatg agcgatgat caagttagat
19461 tagtcagaca cgtttagcat acctgtgtt acgoggtatg ggtatgagtg aagaatgat tcaagaaac
19531 caaaagagtg gcgcatttga gttgttgac aaagatatgg acgttaata cttaacaaa gatgaatg
19601 accaatgat tgagaacctat ttatgtgaa tgcgaagaa tatcatgct ttgtcaaat cagttaactt
19671 taattctgac gaggtttaacg gaaatgtacc tatcattgga atgaacatta aacttatggc tttagagaac
19741 aagtgatga cgtttgacg taagatgaca gctatgtga ggtatcaatt caaagtattt ttatctgcat
19811 taagcgtaa aggttacaac ttggtatgt atagtattt aaacctgata tttaagtcca ctcgtaacat
19881 tccagttaat aagttagaag aatcaacagt gctaataac ctgaagggag aggtttcaga acgaacagg
19951 ttaggacaat cacaactagt tgatgatgtt gattacgaat tagacgaat ggaagagaa agtcttgaat
20021 ttatgacaa attacctgac atagatgag gtgacgaaa tgacaaatcc caaataaacc aatcagatg
20091 atattgtga gtatatgag gggttaact cttaagcaga aaacaaata gaacactat ttgctaagc
20161 acttaagag ataaaacaaa tcatcgaga tatgtttgag aatatcaaa atgatgatgt gtatgtbca
20231 tggactgaat tcaataaata caacaggctc aataaggagt taactcgtat aggtacaatg ttgacttatg
20301 actataggca agtagctaa atgattcaga agtcaacaga agatgcttat atagaaaaat tcttatgag
20371 cctttattta tatgaatgg cgaagtcacac atctatgac tttaggttc cgaagtaaga ggttaacaa
20441 tcaagtattg acaacactat tgaattcatt cgtttaatgc caacactaca aaacatcgt gatgaatg
20511 tcaaaaagat acgtatgcac attacacaag gtattatgag tggagaggtt tacttaaga tagctaaagc
20581 aatcogtat gatgtcgga tgcctaaag tcaatcattg cgtgtggctc gtacagaagc aggcagagca
20651 atgtcacag ctggacttga tagcgcaatg gttgctaaag ataacgggtt gaatatgaag aaaggttggc

5

194

10

15

20

25

30

35

40

45

50

55

20721 atgtctactaa agatcacaga acacgtgata ctcatcgta tttagatggg gaatcagtg aaatagatca
20791 gaatttttaa tcaagtggtt gtgttgggca ggcgcccaag ctatttattg gtgttaaacg tgcgaagag
20861 aatatttaatt gtcgttgcaa attactttat tatatttgat aaaaatgaatt gccnactgta atgagagcac
20931 gtaagagaga tggtaaaat gaagtattcc cattcatgac ttatcgtgag tgggagaat atagcgaaa
21001 aggtggtaat tgatattgat tttaaaataa aagtaaatgt tgatattggc gaagctatag aaaaagtaga
21071 acgattaaa tcccttgtag aagagataat agagttaaca aacyaaaaag ttgtgtgaaa cgttaacgtt
21141 aaaaatgaag ctgatttaga tatgggttaa acatctatta gcgaagaaa tgctaaaaat aatgatttca
21211 cactttttta gtgtctctt tgctactcga ccttagcatg tcgttaaaat gcttttttact atgcactttt
21281 cggactgtta gggtagcaga agggcaaaaa ggaagtttga tatatgaata tcgaagaagt taagtctttt
21351 ttgtgaagac acaagagaga taagaagta aagattatc taaggagact taagacggtg tctgttgatg
21421 acgttaaggg ctttttagat acagaagaag gtaaacgatt cattcaacct gaatttagatc gttatcattc
21491 gaagagatta gaatcatgga aagagaaaaa tcttgaggat ctaatcgaac aagaagtacg gaagcgtaat
21561 cctgagcaat cagaagaaca aaaaactgatt agtgccttgc aacagagatt agaaaaacgc gacgcagagg
21631 caaaaactga gaagttaa gaacacgcgc taggttaagc gcaggaacta aalttaccac catccttagt
21701 tgatagattt ttggcgatt ctgatgaaga tactgagcaa aacttaaaag ctttaaaaga aacctttgac
21771 aagattgttc aaaaagcgtt tgagcttaaa tttaaatcga gtggaagaga tgttaaaaga tcacgaatc
21841 aagatttaga cctttcaaat gtaagatcca ttgaagaagt ggcgaagaa atcaatatta gaanaaag
21911 tgaggttaaa aatatggca actccaact acacgccagg caatgttatt ttatcggatt ttaaaaacgg
21981 cgttatccca gcgaacaag gtactttaat catgaagac attatggcta attcagcaat tatgaataat
22051 gctaaaaatg agccaatgac agcacaagag aaaaatttta cttacttagc aaaaaggtga ggcgctact
22121 gggtagcaga aacggaaggt attcaactt ctaagcctga atatgcgcaa gcgaagaagg aagctaaaga
22191 aattgggtga attattcgt tatcaaaaga gtttcttaaa tggactgcaa aagattttct taatgggtt
22261 aaactcttaa ttgcagaggc attttacaaa ggggttgacc aagctgtatt ctttggtaact aatcacctt
22331 acacacattc aactagtgtt aaacgccttg ttgaagggcg agagagaaa ggttaacgtt ttacagatac
22401 taataattta tacytagaac tttygcatt aatggctact attgaagtg aagagttaga tccaaagga
22471 gtaataacta caagtcatt cagaagtaaa atggttaag ctttagatgc taatgacaga ccattatttg
22541 atgtcaacgg gaacgagatt atgggttacc cactatctta tactggagcg gatgtatag acaaaaagaa
22611 tctgaagatg ccacgttaac gacgttaaca gcactcagatg cttctggcca accagtatca ttattgaac
22681 ggtatagttt cgtttacgt gcgacatgc atattgcata catgaacgtt aaacagaag cgttcgcaac
22751 gcttaaacca actgaatagg agggatattg atggctaatc ctgcagaaga gattaaagta aaaaagaca
22821 atatgactat tactgttaca aagaagcatt ttgactctta ttacagctct gtcggttaca aagaggttaa
22891 atccagctgt actacgtctg ataaagcga gtgataaaaa tgactcttta tgaagatgtt aaactttac
23011 tcaagaaaaa tggagtggaa gttaaaagtg atgaagaaga aatatttaag atgggaagtg aoggaatatt
23101 agaagatggt agggatataa caaacatga ctttatgaag gatggctaac tcaattatcc ttactcaatc
23171 aaaaagtagt tgcgagatgt cctagagat tatcaacgac ctgaagttaa aagaatttta aagtcaagaa
23241 atgtggggac agtgcgtac acttataacg atgtgtctcc tgattacatt agtggagat taacaggtta
23311 taaacgagca aagttctatc cgtttaaac aataaggtat aggtgttgtt tgtgtttaa ccatagcag
23381 atctccca cactatttct attggaagta tcaaaaaggt agggagat ccaattatcc aagagcgctt
23451 tgaacagat aaaaacatta aaggtttat ggaacgctt actacatctg aaaaactaaa atttcatcaa
23521 atgtcacaag aatattgacg aaacctatat gtaccttatg acttgccaat atcttaaaac aatttatttg
23591 agtatgaggg tagaattctt agtattgaag gtgattctgt agatcagggc ggcacacatg aatattagtt
23661 actacgactt aagcaggtgc catatggcaa aagtttaagta cgtgtgtgat agcatgtgtg tgaatttga
23731 taagtctgat aaaaaaatg aagagtggtt taaaaaaggt attgctaaaa caacgagaa gatttcaac
23801 actgcttag cattagctcc tgttgactta ggttttttag aagaagatg tgactttaaa tattctgatg
23871 gtgggttato caggtgtata agtgcgtgcg cagattatgc aatacagctt gaatacggta ctggtatata
23941 tgcactcgtt cctggtggtta gtgcgtctac aaagattcgc tggagtttta aaggtgatga cggcgaatgg
24011 tacaccacat atggtcaagc gccacagcca ttttggaacc ctgcaattga cgcagagcgc aagacattcg
24081 agcagtattt ttcatagagg tggttaaata tgtgggtatc agttgagcct gaacttaca atcaaatata
24151 taaaagatta atctcagacc ctaacattaa caaactagtt gatgataggg tttttgactg tgtcaagat
24221 gacgctgttt ncccatatat tgtgtgggtt gnatcaacg tcaactaaca cgaacttagc gcaacattga
24291 gagaacagct cgttatgtgc atacatggtt attcagagtt cgttacaaca tacyagggta agctcatttt
24361 aagcgcgata ggttatgtgc ttaacagacc tatagaata gataattac agtttcaatt tgcctgatac
24431 gatagtcag cagttatccc tgatataagc aggtttacta agcatggcga gatagcgctt ttatttaagt
24501 acagacataa aagaanaaac gaaggagtgt attaaatggc gcaaaaaaac tatttagcag ttgtacgtcc
24571 agctgaacct gacttagatc cagtagaatc ttattattta gctgacttac aagaaggtgg acatacgtt
24641 aaaaatgatt tagctgaat agtaagaggc ggtaaaaagc actattctcc caatgcgaatg tcgaatcat
24711 ttaaatatc aattgttaat gtgcctggag ataaaggaat tgaagcagtg aaacacagctg taocaaacag
24781 tggcagttg cgtatatggc ttatgagcgt taataacgt gcagacggtta aacatcacgg aatgtttggt
24851 tatgtgttc cagaatcatt tgaatgtca tttgatgtg aagtgacaa aatcgaacta tcaataaag
24921 ttaaatgaa tacagcagaa ggtgctgaag ataacttgc gaaagagtg tttgaagctg caggtgccc
24991 tcaagtga taagaaaaat ccgctgaata agtcggaaca ttcgagaatc aaaaagaagc tagtgttga
25061 tctgattcac acacggaaga ccatctatg taaactaata gatcaagggg gcgttaagctc cctattttt
25131 tataaaaaa ttgaanaag gtatataatt tgactgaatt taactccaat acaacttaa aatattatga
25201 cgggaaaaa gattacgaag tagaagcaaa agtaacattt gcaattgacc gaaaagctga aaaaattcca
25271 gaagatagcg aagatgggag aaaaaggaga atgcagagat tcaatgttat ctttaacggt ttgctagaat
25341 ctgaaacaa agcgtattta caattttggg aatgtgctac tgcattatta aaaaaccac caactcga
25411 acattagaa aagcaattg atgattttat cactgaanaa gaggatattt tgccttattt acaaggggct
25481 ttgacaaac ttaacaaatg tggtttttct aagagggaga gtgcctcga ctggagaca ttgaacaaag
25551 caccgaatat ggcacaaagc gaggacaaag aatgacgaa agcagggata aatagatga aagagatta
25621 caaggaatc atgggagcag aacttaccac gattactcaa aataagggca actgacagct agatatttag
25691 gatatacccc tgaacatgaa ttgttagcac taacacctgc tgaatggcgt gattggctta ttggtggca
25761 ggaaggtac ctgattcaa gacaattatt aattgaacaa gcgcaagcta acggttagt acaagcttct
25831 aagaggttaa ctgattgat tctgacatt gagaacacac gttacgaaat aagagaacct ggtagctatg
25901 ctggtgaca aaaaactga ttgaagaag aaaaagaag acgtgaactc ttcaagaag gtacagaaga
25971 attccttga tcgaaggag gtagcctt gtagctctt ttattggcaa agattatggc caatattaga

5

195

10

15

20

25

30

35

40

45

50

26041 gatttccaaa gcaacgttaag gaaagctcaa cgtattagcaa agacgtctgt accaaacgaa attgaanacg
26111 atgtaaaagc agatatttca agattccaaa gagctttaca acgcgtctaa tcaatggctc aacgatggcg
26181 agagcattct gttaaattat tcatgaaaac agatgagtat aaagcgaatt tagaacggcg taaggtccaa
26251 gttagcgcat taaacaaca taagtagat ttgaaactaa gtaacactga attaatggcc aaatataatg
26321 caactaaagc tactgtcgaa gcttggagaa aacatgttgt taagtggat tagargcaa acccgctcaa
26391 aatggcggtt aaaggggtta aagaagattt aatagatctt agcagggcata gttttgatat tgattccagc
26461 agatggaaat taggaataaa attcacaana gaattcaatg aagtcgaagg agcagttaaa cgttctttcg
26531 gaagaattgg tcagattatg agaaaaaag taatgggaac aagtatattt tggggtaaac ttaacaaactc
26601 attgaanagc tacggcgaga aaatggacgc cttagctact aaatccgaa ctttcggtac tatcttcg
26671 caacaggtca aaggtctaat gattgctagt atacaagcat tgataccagt gattgcccga ttagtacctg
26741 caataatggc agtactcaat goggttggtg tattaggttg tgggtttta ggtttagtgt gogcattctc
26811 tttccaggtt cttggagttg ttggctttgg tgcaatggct attagcgctc ttaaaatggt tgaagatgga
26881 acattggcag taacaaaaga agttcaaaac tttagagatg cgagcgatca gttaaaaact acatggogtg
26951 atattgttaa agagaataca gcaagtatct ttaatgcgat gtcagcaggt atcagagggc ttacaagtrc
27021 gatgtctcaa ttaaaacctt tottatccga agtatctatg ctagtgtgaag caaacgcagc cgagtttga
27091 aatgggttta aacattccga aacagctaaag aaagcgtttg aagcattgaa tagcataggt ggcgaactct
27161 tcggagattt attgaacgct gcaggagatg ttggcgacgg attagttaac attttcactc aattatgctc
27231 gttgttcaaa ttgtgtcttc aaggactaca gaacatgtct atagctttcc aaatgtggcg taatagtgt
27301 gctgctcaga atgctattaa agcgtttatt gactacacta ccactaactt acctaaagatt ggtcagatrt
27371 ttgttaatgt gttcgtggtt ntggtaattt taatgattgc ttttgacaaa aacagtcca acatttttga
27441 ttgtttggtt aaattaaact ctcaatttag agcatggtca gaacaaagta gacaaactaca agggttttaa
27511 gactttatca gttatgttca agagaatggt cctactatta tgcagttaat cggtaatatc gtaaaagcat
27581 tagttcgttt tggatctaga atggctctta tagcttagta attgttagac ttatcacta cttaagctgg
27651 attatcgctt aaactattcg aaacacccc agctatagca caagtgtgtg gogttatggg tattttaggc
27721 ggtgtatttt gggcttcaat ggctccgatt ttgtgtataa tagctgtact tacaatagtg tttggttga
27791 gctattcag cgtcaactga agatatttag acttcgttag aacatcaagt ttagtactg gagctacgga
27861 agcatataa ggtgactatg gttcgtattc agcaactatt tttagcattg ttgcagtaat tgggtcattc
27931 attgggtctc tctgttattt atggaaaaa aacgaagaa ttgaataac tattactgaa gogtggaaac
28001 ggtgtaaaac ggcagtctct ggtgcgattc aaggtgtagt cggctggtta actgaattg tgggcaaat
28071 ccaactcacc ttacaaccca taatgctcat attgcagta tttaggcaaa cattcatgca agttttaggt
28141 gttttgtgaa taggcacatc tcaaacgctt atgaatatca tacaaggttt gggacttta attcaattg
28211 cgttccaagc cataggaaca gtgatatccg tagcagctca aatcatagta ggtttgttca ctgctttaa
28281 tcagtgtctt actggcgact tctcaggtgc ttgggaagact attaaaacta cggttaccaaa tgtgctgat
28351 aogatttggc aatcatgca atcagtttgg gagtcaatta tgggtttttt aactggogta atgaatgaa
28421 acttcttat gtttggtaaa agttgttca agtatggag tacaactact aattttgtta gcagtattg
28491 gaacactggt acaagttggt tcagtgcagt ggttgcagt gtgagtgaaa aaatggggca agcactaaac
28561 ttattatca caaaaggttc tgaatgggtt tcaaacattt ggaatacagt tacaagtttc gogagtaaa
28631 tagctgatgt gtttaaaaaa gttgtctcaa atgtatggta cggatagat gatgcactg gtaagattaa
28701 actttcttct agtacttct taatgcggc agcggaaata atcggcaaa tagctgaggg ttagccaaa
28771 tctgcgcaca aagtagtcag cgggtaggc gatgogatt catcagcttg ggaactctga acttcattc
28841 taagtggaac cggtagaggt agtagctag gtaaaaggtt agcgttatca caagcaaaag taattgact
28911 agactttggc agtgctctta ataaagagct atcctctact ttgacagata gtatagtaaa tccctgaat
28981 acttctatag acagacacat gactagcgat gttcaacata gcttaaaaga aaataataga cctattgtga
29051 atgaacagat tagaaatgag ggcagccttg atttaattaa atcagcatt gatgacatga acgctataga
29121 cggagatttc aactattat aagggaggtt ttttagtga tagcgacga tatagaatga ataaagaaatg
29191 gttcacagta tgcgctcagt gacaactctt tcaattataa tcaacttgaa gtatgtgaat ataacgttac
29261 agggcagga tatcatgta actattctga tatagaggtt attgatgga gatttcataa ttacgttaa
29331 gaagaactta aaaaagtaga gtttaagata aggtataaag taactaaat tgcattatgt tcacatttaa
29401 agtcagagct ccaagcacta ttgtctggac gtttttattt aagggaaata gctacaccag acaattcaat
29471 taagtatgag catatattag atatacaca agacaaacaa gcatttgagc ttgattatgt ttagggagca
29541 caactttttg taggactagt aagtgaaagt tcttttgaca caacacaaac atcaggggaa tttctttgt
29611 cgtttgaac aacgaacta ccatactttg aagtggtcgg ttatagtact gatcttgaaa gtaataaaga
29681 ccttgaaaaa tggctcggtac ctgatatgatt gcttacaac gaaggtgata agaggcgta aatgacattt
29751 tacaacacta actcaggaga agtttattat aacggtgatg ttcttttaac acagtttaat cagtttaatg
29821 ttgtgaanaa agagttagct gaagatgta aagotaatga taaggatgga ttactttct atacagataa
29891 aggaatatct tcagttatta aggaagtga tttaaaagcc ggagataaaa taactctoga cggtaaacat
29961 acctatagag gttatttaaa tatagattct ttttaaaaa ctttagaaca accggtttta tatccaggct
30031 ggaatcgat caagtctaat aaagtatga acaaatatc atttagacac aaattatatt ttatagtaag
30101 agtgcctat gccattttta ttaaaaagtc taccgggtgt agggcagctt attaatgta gtacaaaggt
30171 aagtaaaaaa ctaaatgaag atagttcttt ggtactaact attatogaga accggagatc gtttgacga
30241 ataggttgta taactaaat gtcgacgac actcatgttg aagtgaaaga tgatttcaac gaattatgaa
30311 ttgtctatct tgataagctt actattggcg aaaaaataag gcttgatc acagctaggc aaaaagact
30381 tgtagacctt acaattctta ggaatttaca agagtataac gaagtttta cggcggttga gttcttcaat
30451 actgtcttta aaggaacggg ttataagatg gttattacat caaaagtaga tgcattcaaa ttcgagggat
30521 taggcacagg agatacacga ttagaatct ttaaaaaagg acttgagcgt tatcctctcg antatgaata
30591 cgtatgcaag actaaaagct ttcatgtga tgatgaatta tctaatgttg ccaattatta cattaaagct
30661 ggtgtgaatg ctgataacgt caaaatacaa gaagatgcat cttaagttra tactttattt aaaggttatg
30731 gttgatttga tggacaacag acttttgca agcgggact acaaatgaa ttcactcacc cattagaca
30801 attgataggt aaaaagaaag cgcacccgct tttgtatgga cgtattaaaa aagaagatag tttaaaaaa
30871 gcaatggagt tattgaana gaaaagtgtc actgcttcta tttcttga ctttgtagog ttactgaaac
30941 atttccaga agctaaccct aaaaaggtg atgttgttag agtgggtgat ttgtcatag gatataaga
31011 cttagtgaga atagtcaaaa tcaactacac tagagatgag tacaatgata tccatagca agatgtatga
31081 tttagagact tcaacaggcg taatogttat acaaaagcag ttcactgatg tgcnaattat gttaaaaagc
31151 taatactcac aaaaatcgac ccatctaaag aactaaaagc attaaagca aaggttaacg caagtttatc
31221 taataatgat gaattggta agcagaatga aaaaataaac gctaaagctg ataatgaa tactaaaaa
31291 gttacaactg ctaaggtac gatcatgac gacttttcta gtcactaag tataagaac accaaactca

55

5

196

10

15

20

25

30

35

40

45

50

55

11361 ttggaaacgat tggcgactct gtatctagag ggtcgacgc aaaaactaat ttcacagaaa tgttaggcaa
11431 gaatttgaaa gctaaaacga ctaactctgc aagaggtggc gcaacaatgg caacagttcc aataggtraa
11501 gaagcggtag aaaaacgcat ttatagacaa gcagagcmaa taagaggaga cctaatcata ttacaaaggca
11571 ctgatgatga ctggttacac ggttattggg caggcgtracc gataggcact gatamaacgg atacaaaacac
11641 gttttacggt gccctttggt ctgcaattga agttattaga aagaataatc cagattcaaa aatctagtgc
11711 atgacagcta caagacaatg ccttatgagt ggtacaacaa tarcgctaa agacacggac aaaaacaaac
11781 taggggtaac acttgaggac tatgtaaacg ctcaaatatt agctttagt gagttagatg taccagtgtt
11851 tgacgcatac cacacagatt actttaagcc atacaatcca gcttttagga aagcagcatg gaggagcggc
11921 ttacacccca acganaaagg tcacgaggtt attatgtacg agttaatcaa ggattattac agtttttacg
11991 actaaaggag gcaaccaatg gcttacggat taattacaag ttacattca atgacaggtc ggaataatgt
12061 tgctcaacat gattataact atcgcttgtt agatgaaggt atgagcaaac ttgagaaatc gtttatatac
12131 catcaaaaag aagaataata cgcacactca cgcgaacaaa tcaaatctt gaatgcagt gttgaagatt
12201 attcaacgta ttttaaatgc cgttttagca atatgattct aggcataaac ggcagcggta tcaatgaagt
12271 aaaaagcgcg cgtattgata atacaggtta tggctatag acattgcaag atcggtttga tcatgattat
12341 tcaacactag atgctttcac taaaagggtt gagaagagctg tagatgaaca ctataagaa tatcagagga
12411 cagaatacgg attcgaaoca aagagcgaag aacgggaatt tatcactgat ttatcgccat atacaaatgc
12481 agtaatgcac tcattttggg tagacccatg aacgaanaat atttatatga cgcagcctcg tccaggtaat
12551 catctacatg tatctagatt gaagcccaac ggacaattta ttgatagatt cgttgttana aacggcggtc
12621 acggtacaca caatcgctat agatadattg atggagaatt atggtattat tcaagctgat tggacagtaa
12691 caaaaacacac aagcttgtag gtttccata tagaactgga gaataaactt atggttaatga atgcgaagt
12761 gtcgtgocga atatatattaa cgcagagatat acgtcagcga ttataatcc tatagaanaa ttaagtattt
12831 tcagagcgtc atataaagct tctgaaagac aagctaaaga ttcatgaa ttcatgaag taagagtgcc
12901 tgacgataat gataaaggta tagcaaaagt atgtatcaa atggatatc ctatggata cacttcagat
12971 acacaaccta tgcnaaggtat caacttatgt caggttatct tatattgta tcaaggtgat tcaatagat
13041 ccaaccccaa tcaactacaa ggtttogata taaaacaaa agaattgta ttttaagac gtatogtat
13111 tggcggtgtg aataataact ttaaggaga cttccaagaa gctgaggtc tagatagta ttacatcta
13181 gaacacggag gtaaacgact ttaaatagg gtactattg gacttggtaa taacagacat cactcaattt
13251 attctatcgg ccaaaagggt gtttaacaa tcttaaaaaa cacttgcact caagttatga tgaagtattc
13321 aggtggaggt gtaaacaggt taccataca gaacccagca tatcttaagt atattacgga agttggtcat
13391 tacttatctc atacgcaaga cacacaaat gcattagatt tccgttacc gaaagcgtt agagatgcag
13461 ggtggtttct ggtatgtact cctggacact ataatgggt tctaagaca gtaactacca gaacagcac
13531 aggcagaatc atgcttaaat tgaacaggt cattgacatt tccaataaga aaaaacacgg agcatggaat
13601 tctctgcgcg aaaaacgggt ttattgggaa catatcccta agagtattac aaaaattaca gatttaaaaa
13671 tctgtgtgtt agattttcat atcactatgt aagaatcaaa acgatttact gattttccta aagactttaa
13741 aggtattgca ggttgatgat tagaagtaaa atogaataca ccaggttaaca caacacaagt attaagaagt
13811 aataacttcc cgtotgcaca tcaattttta gttagaacct ttggtactgg tggcgttgg aaatggagtt
13881 tctatgaagg aaggtgggtt gaataatgt agtagataa ttttogaag acgtaactt aatcgagta
13951 caaacaacat cacaataata tccaattatt gacacaaaca tcaagttcta tgaatcagat agaggaactg
14021 ggtttttaa ttttgcagta actaagaata acagacggt atctataagt tctgaacatg ttaaacatc
14091 tatcgtgtta aaaaacgatg attataacgt agtagaggt gcttatatt cagacgaatt aacgataga
14161 gcaacaatta atggcggtt gcagtatgt ataccgaatg aattttttaa acatcaggc aaggtgcatg
14231 ctacagcatt ctttacacaa aacgggagta ataatgtgt tgttgaacgt caatttagct tcaatttga
14301 aatatgatta gttagtgggt ttgatggtat acaaaagct ttttatatca aatctattca agatactatc
14371 gaagcagctg gtaaaagact taaccaatla aagcaagata tggatgatac acaaacgtta atagcaaaag
14441 tgaatgatag tgcgacaaa ggcatcaca aatogaat caagcaaaac gaagctatca agactattac
14511 tgcgacgcaa actagtgcaa cacaagctgt tacagctgaa gtcgataaaa tagttgaaa agagcaagcg
14581 attttgaac gtgttaacga agttgaacaa ccaatcaatg gcgctgacct tgttaaggt aactcaacaa
14651 caaattggca aaggtctaaa cttacagatg attacggtaa agcaattgaa togtatgagc agtccataga
14721 tagcgtttta agcgagttta acacatctag gattattcat attactaatg caacagatgc gccagaaaag
14791 acggatatag gacggttaga gaagcctgga caagatgggt ttgatgagc ttotcgttc gatgaatcaa
14861 cttatacatc aagcaaatct ggtgtgttag ttgtttatgt tgttgataat aactctgctc gtcgaacatg
14931 gaccccgac gattcaaacg atgagtacac aaaaacaaa atctacggca catggtaccg gttttatata
15001 aaagatgatg gaacttaac taagcaattt gttgaagaaa cgtctaaaca cgttttaaat caagctaagc
15071 agtatgtaga tgataaatc ggaacaaaga gctggcaaca acataagatg acagaggcga atggtcaatc
15141 aattcaagtt aacttaata atgocaaagg cgaatttggga tatttaactg ctggttaatta ctatgcaac
15211 agagtgccgg atttaccagg tagtgttgaa agttatgagg gttattttatc ggtattcgtt aagacgata
15281 caaacagct atttaactc acgcttata acttaaaaa gatttacaca cgaatcaatc caaacggcag
15351 acttgacaa cagtgagcag ttccaatga acataagta aoggtattgt tgcaggtggy ngcaaatggt
15421 taggtacaaa caatcaatct aacogaacca tacacaaact attctattt attagtaagt ggaacttatc
15491 caggtggcgt tattgagga ttccgactaa ccaattacc taatgcaatt caatcaagta aagogaatgt
15561 agttgactca gacgttaacg gtggcggtat ttatgaggtc ttactatcca aaacagtag cactactta
15631 agaatcgata acgagtgtga ctttgattta ggttaaacat caggttctcg agogaatgcc aacaaagta
15701 ctataactaa aatttgggg tggaaataat gaaaatcaca gtaaatgata aaaaagagt tatcggatc
15771 gtttaactg cgggtttacg caatagtta gatgtagac atacaatgt gtcatacaa ttcaagaag
15841 atttcgaact tagaagttc gttttcacta acggcgaat taaatacaat agcaatttcg aaaaagaga
15911 cgtacgcaat gcatcaaac acaaatgtc gtcagattta agtgatgag aacttcggg aatggttga
15981 agtatgcana tgcagatgac gcaagtgaac atgttgacaa tgcattgac gcaacaaac gctatgttaa
16051 cacaacaggt gacgcaactg aaaaactaca aaaaataac tgaaggggac gtttaaatga tgaagtgat
16121 ttatccaact tttaagaca taaaacttt ttatgtgtg ggttgcata aaaaagca aattagtg
16191 taogtagaca tgggtgtat cgaacaaaga gaatatgcat tgatcaatg tgaanaaat ccaagggaa
16261 aagatgaana gtcacaggt taatgttga ggtttttta ttacacaa agtaggtggc gtaatgttg
16331 gatttaacca acggcagaa catgaatggc gaattagaag attagaag aatgataaa caatgcttag
16401 caactccaat gagattaaat taggtcaaaa aactcaagc caagttaaca tttaattaga taaaactta
16471 gatgtatcc agagggaaag acagatgac gaaaaaata agaaagaaa cgaacaaat atacgcata
16541 tgaataatgt gattctcgt tgcagagga ctacttcag tgcgtttgc atagctttac taagaactat
16611 ttttgggtat taaaggaggt gattaccatg cttaaagga ttttagata tagctttcgg cgtgcttct

5

197

10

15

20

25

30

35

40

45

50

55

36681 ggtttggttaa atgtaaataa cagttaaagag tcagtgcttc ggcactggct tttttatttg attgaaatga
 36751 ggtgcataca tgggtattacc taacccaaag actagaaagc ctacagctag tgaagtgggtg gagtgggcaa
 36821 agtcgaatat tggtaagagag attaatatag ataattatcg gggcagctcaa tggctgggata caccctaactt
 36891 tatttttaaa agatatttggg gttttgtaac atggggcaat gctaaaggata tggctaattta cagatatcct
 36961 aagggtttcc gattctatcg ttattcatct ggatttgtac cggaaacctgg agacatcgcn gtttggcacc
 37031 ctggcaacgg aataggttcg gacggacaca cggcaatagt agtaggacca tctaataaaa gttattttta
 37101 tagcgttgac caaaactggg ttaattctaa tagtgggaca ggttctccag gaagattagt aagacacctt
 37171 tatgtaaagc ttacagggtt tgttagggct ccatactcaa aagatactag caaacctagt agtactgata
 37241 caagtccagc atcaaaagc natgactcaa caattactgg cgaagcgaaag aaacccgaat ttaaaagag
 37311 taanaacagta aaatacactg cttacagcaa tgttttagat aaagaagagc acttcattga tcatatagtt
 37381 gtaattgggtg atgaacgctc agatattcaa ggattatata taanaagatc aatgcattcg cgttctgtag
 37451 acgaactgta tacgaaaga aataagttta taagcgatta tgaatacccg catttatatg tcatagaga
 37521 ggctacatgg cttgctagac caaccaattt tgaatgacgg cgtcaccctta atgggttagt tattgaagta
 37591 tctggtggct aaacagatag caaacgacaa ttcttattga atcaaatata agcgttaata cgtggtgttt
 37661 ggttattgtc agggattgat aaaaacttat ctgaaacgac gttaaaggta gaccttaata tttgggttag
 37731 catgaagat ttaatttaatt acgacttgat taagcaaggt ataccggata acgcaaaagta tgaagcaagt
 37801 aaaaagaaaa tgcctgagac atacattaaa cagatataat tgaacggaga aaatataaaa gaagtaacga
 37871 caaaaacaa aataagaatt agtgataaaa catcagttga cagtgggtcc acagaggcc ctactccatc
 37941 agacgaaaaa ccaagcatcg ttactgaaac aagctcattc acattccagc agcactgga tagacaaatg
 38011 tctagggtta acccgaaaaa atctcataca tggggctggg ctaatgcaac acgagcaca acyagctcgg
 38081 caatgaatgt taaggaataa tgggaagata acacgcaatg ctatcaaatg ctttaattag gcaagatca
 38151 aggcatttca gttagtgcgc ttaacaaat acttaaaaga aaaggaacgc tcgaacggaca aggcacagca
 38221 ttcgctgaag cttgtgaaga aaacaacatt aacgaaattt atttgatcgc gcaogcttcc ttgaaagtgt
 38291 gatacgggac agtaacttc gctagtggta gatacgggtc atataattac ttccggcaca agcaatcatg
 38361 caaogccctt gattatgcaa tgaogtttgc taanaaaaa ggttggacat ctcccgcaaa agcaatcatg
 38431 ggggttgcta gcttctgaag aaaggttaac atcaataaag gtcaaaacac atttgaccga attagatgga
 38501 atcctaagaa tccagctacc caccatccg ctactgctat agagtgggtc caactcaag caagtaaat
 38571 cyctcaagta tataaacaaa tgggtttaa aggtatctac ttcaacaggg ataatataaa ataaagaggt
 38641 gtcgaatgt acaaaataaa agatgtgaa acgagaataa aaaaatgctg tgttgactta ggtgacattg
 38711 gctgtcgatt ctacactgaa gatgaataa cagcatctat aagaatgggt atcaatgaca aacaaggtcg
 38781 tatcgatcta aaagcacatg gcttaaaccc tagattacat ttgtttatgg aagatggctc tatattcaaa
 38851 aatagagcccc ttattatcga cgttgttga aaaggttccc ttacctcaa aatacctaaa aaggttatca
 38921 aaacagctgg ttatgctcgc tgaagctgt ttttagagaa agaaagaaa aaaaatacatg tgcnaaactt
 38991 tcttctcaat atcgttgata gtggtattga atctgctgta gcaaaagaaa tcatgttaa attggttagat
 39061 atcgtattta ogagaatttt aaagataac gogacagatt tattgagcaa agactttaa gaaaaaatg
 39131 atcaagatgt catttttao atcgaaagaa atgaagtag atttaaggt gcaaaaggtg ataaagggga
 39201 atcgagacaa cctgggtgga aggtgatac aggttaaaaaa ggagaaacag gcgcacccgg taanaacgtt
 39271 actgtagat catcaatcc tgccactaaa atgtggcnaa ttgatggtaa agatacagat atcaaaagcag
 39341 aactagatt ttgggacaaa atcaatctg caaatgttga aggttagaaa gataaattgc aagaagttaa
 39411 aaatatcaaa gatcaaacct tcaacgactc taanaagat aoggtattca aatttgctga actagttagat
 39481 agcgcgctg aatctatgaa tacattaaaga gaattagcag aagcaataca aaacactct atttcagaaa
 39551 gttgattgca acagattggc tcaaaagta gtacagaaga ttttgaggaa ttcaaaacaa cactaaacga
 39621 ttatatgct ccaaaaatc ataactcga tgaagcgtat gttttgtcat ctcaagcttt tactaaacaa
 39691 caagcggata attttatca actaaaagc gcatctcaac cgaaggttaa aatttggaac ggaacagaaa
 39761 atgaatataa ctatataat caaaaagacc caatacact ttacttaatt aaggggtgat ttttatggaa
 39831 ggttaattta aaatgttaa gaagtattat tacyaaggtg aagaatatac aaagtatat gctggaataa
 39901 tcaagtatg gaaaaagcct tcatctttg taataaaacc cttaactaaa aataaatatc cggatagcat
 39971 aagaataca acagcaaaat ggacaatnaa tggagttaa cctaataaa gttatcaggt gacaaatgaa
 40041 aatgtacgta gcggtataat gaggttttcg caaactaatt taggttcaag tgatttagga atatcaggag
 40111 tcaatagcgg agttgcaagt aaaaattca actttagtaa tcttccagg atgtgtgat tcaataaag
 40181 tgaattttat tcaggatctc caacattgac cattgaataa ttttaacga ctaatttttt agtcgttttt
 40251 tattttggtt aaagggagca acaaaatgga tgcanaagta atcaacagat acatcgatt gatcttagca
 40321 ttagttaact aattcttagc gaacaaaggt attagccga tccagtaga cgaatgagat atatcatcaa
 40391 taactactac tgttgttgc ttatatacta cgtataaaga caatocaa ca tctcaagag gtaaatgggc
 40461 aaatcaaaag ctaaaagaa ataaagctga aaacaagtat agnaaaagca cagggcaagc gccaatcaa
 40531 gaagtaatga cactacgaa tatgaacgac acaaatgatt taggttaggt gttgaccaat gttgatca
 40601 aaaaacaaag cagaaaaatg gtttgataat tcaataggga agcagttca tctgtattg tttatggat
 40671 ttcagtgtta cgaatcagca aatattgttt ttatgatagc aacaggcgaa aggttacaa gtttatagc
 40741 ttataatatt ccaattgata ataaagcaag gattgaaaaa tccggcgaaa taattaaaaa ctatgatac
 40811 ttttaccgc aaagttgga tttgtcgtt tcccgtaa agtattgtg cggagctgga catgttgaaa
 40881 ttgttgagag cgaatattta aacactttca catcatatg gcaaaattg aatggtaaa gttggacaaa
 40951 tggcgttgcc caactcgtt ggggtccga aactgttaca agacatgtt attattacga tgaaccaaag
 41021 tattttatta gattaaattt cccagataaa gtaagtgtt gagataaagc taagaagcgtt ataaagcaa
 41091 caactgcaa aaagcaagca gaatttaaac ctaaaaaaat tatgcttga gccggtcatg gttataacga
 41161 tcttgagca gtaagaaacg gaacaaacga acgcatattt atcgttaaat atataacgc aaatatcgt
 41231 aagtatttaa gacatgagc tcatgaagt gcatatattg gtggtcaag tcaatcaaa gacatgatac
 41301 aagatactgc atcgggtgtt atgttaggaa atataaaga tcatgatta tattgggtta aatcacaggg
 41371 gcatgacatt gttctagaga ttcatataga cgcagcagga gaaaatgcaa gtcgtggga tgttatatc
 41441 tcaagtcaat tcaatgcgga tactattgat aaagtatatc aagatgttat taanaatac traggacaaa
 41511 taagaggtgt aacactcgt natgatttac tgaaggttaa tgcatacga gaaataaata tcaattatcg
 41581 tttactgaa ttagggttta ttaataaaa aaagatatg gattggatta agagaatba tgaattgtat
 41651 tctaaattaa tagctgtgc gattcatggt aagcctatag gttggtttgt agctgttaat gttaaaacat
 41721 cagctaaaaa ccaaaaaat ccacacagtc cagcaggtta tacaattgat aagaataag tgcctataa
 41791 aaaaagact ggttaattca cagttgctaa tgttaaggt aataacgtaa ggaacggcta ttaactaat
 41861 tcaagaaata caggtgtatt acctataac gcaacactc aatattgag cgaatattgc atcaatgggt
 41931 atagatggat taattatatt gctaagtgt gacaacgtcg ctatttgag acagagagag tagataaagc

5

198

10

15

20

25

30

35

40

45

50

55

```
42001 aggtaatagg ataagtagtt ttggttaagtt tagcacgatt tagtatttac ttagaataaa aatttttgcta
42071 catlaattat agggaaatctt acagttatla aataactatt tggatggatg tcaatattcc, tatacacttt
42141 ttaacattac tctcaagatt taagttaga taacaggcag gtactacggt acttgccat ttttttgta
42211 taagttaatt acattaccag taaccaatct ggcttaaaac cacatttccg gtaggcaatc cggctatgca
42281 gaggacttac ttgcgttaag tagtaagaag ctgactgcat attttaaacc cccatactag ttgctgggtg
42351 gttgtttttt atgttatatt ataaatgac aaaccacacc acctattaat ttaggagtggt ggttattttt
42421 tatcaaaaa aaacgaanaa aagttcataa aagttattgc atatcacgtt taaccgtggt atantaaggt
42491 ataccagtg agaggaggat aaaaagtgtt agaaaatttt aaaaactatag cagaaaatgc ctttatata
42561 atgtcagcaa ttgcatagc gaaaaacttg aaaaagacg ataagtaagt agacaagccc gaaagggtg
42631 tctatatata aattctaaca ctaaaatact atgaaaacaa ttacattat tttatcatt cttaattgg
42701 taacgttgtt tttaggcaac gatataagta aaagtgttgt tgcactgctt actactttac tgcctatcaa
42771 tttatggag agggataaaa atgacagcaa taagaagaat attgaaatca atagaaaagt tattcgaaaa
42841 agaaacggga tatcaaatg ctcaaaatc cggattacca tatcaactg tgcagagatt aagaaatgga
42911 aaaaactctt tatcagatgc cagatttaga acgataataa agttatacga gtatcaaga tgccttgaaa
42981 acgagaaga taataaaaag gagccaaaaa tatgtttgtt acaaaagag aattttaaac tttgagtga
43051 aaagaaatgt ttgaatcagg taaaaacttt ataaaaatta cagatggag acatgcaata tattgggtga
43121 atgatagata cgtagtactt gaccataaaa aaggygattt gtacccgcaa aaagcatacc caaatatat
43191 caaaagaaaa ttagtaagtt aaataaattag aaaaaccag- ctttaattgac gtggttatct tttagggttg
43261 cgcgtgtcaa atacgtgtca atttagttct atttctttag tttctttct aaacttaatt gcttgtaaac
43331 cgcattgta taggcttttc agctatatac caagataaga tttatccgc cgtctccata aaaaatagct
43401 tggaaacctt gatttaattg ggttttaac tagcaagtg caaatatgt tcaagaaaat aactttctga
43471 cccgttgacc ttgtctttt ttatgttcat caagtaagtg agagtagggt tctaaagta tagatatatt
43541 ataattggctt aatcttttgc taatatattc aatagg
```

Table 10

Bacteriophage 96 ORFs list

STD	LAN	FRA	POS	a.a.	RBS sequence	STA	STO
100733	96ORF001	1	25999..29142	1047	ccttgaatcgaaaggaggttagcct	ttg	taa
100734	96ORF002	1	32008..33906	632	tttttaacgactaaaggaggaaccca	atg	taa
100735	96ORF003	1	30109..31995	628	tttatatttttgatgaaggagtagcct	atg	taa
100736	96ORF004	1	36760..38634	624	attttgattgaaatgaggtgcatac	atg	taa
100737	96ORF005	3	33903..35729	608	gtttattcgaaaggaggtggttga	ata	taa
100738	96ORF006	2	40589..42043	484	aatgatttagggtaggtgttgacca	atg	tga
100739	96ORF007	1	18652..20091	479	tatacacacatactaaacctgaacg	att	tga
100740	96ORF008	2	8960..10201	413	tggcagattttggggcgataaaccg	atg	tga
100741	96ORF009	2	17447..18670	407	gacgcaataacgggaagtgtcylca	atg	tga
100742	96ORF010	1	38647..39819	390	taaaataaaataaaggaggtgtgtaa	atg	tga
100743	96ORF011	-3	119..1195	358	gtagctcgccacccttattatttt	ttg	tga
100744	96ORF012	2	20045..21013	322	tttaacgacaaattaccctgacatag	atg	tga
100745	96ORF013	3	29157..30098	313	acttattataaggagggtttgttag	ttg	tga
100746	96ORF014	1	21935..22839	304	agaaaataaagtgggttaataaant	atg	cag
100747	96ORF015	1	5812..6591	259	atcacaggttaaaaggtgggaataag	atg	tga
100748	96ORF016	1	7852..8607	251	aataaaattgtgaaggaggaagaaa	atg	tga
100749	96ORF017	3	3444..4190	248	aattttaacattaatatccatttaa	gtg	tga
100750	96ORF018	-3	28281..29000	239	taagctatgttgaaacatcgctagtc	atg	tga
100751	96ORF019	3	7188..7859	223	tttacggttcttaggacgtgtttaa	atg	tga
100752	96ORF020	3	21324..21908	194	gaaggggcaaaaggaggttttgatat	atg	tga
100753	96ORF021	3	6612..7175	187	attaaaaatttaattaaaaggaggtt	ata	tga
100754	96ORF022	2	24536..25093	185	aaagaaaaacgaaggtgtattaa	atg	tga
100755	96ORF023	1	5275..5811	178	catgaaattcgtaggaggtatgaaaa	gtg	tga
100756	96ORF024	3	14481..15014	177	taaaacgattaggagatcaacgataa	atg	tga
100757	96ORF025	2	25157..25666	169	ataaaaaaattgaaaggaggtatat	att	tga
100758	96ORF026	-3	15084..15590	168	tcattcttaacatagcccttaattc	atg	tga
100759	96ORF027	-1	1229..1732	167	aatagcaaataaagggtgttaaac	atg	tga
100760	96ORF028	1	16960..17454	164	aaggcgctgtgatacagtgaaaaaa	ttg	tga
100761	96ORF029	-1	1736..2227	163	catgagaaaaggaggtcatataaaag	atg	tga
100762	96ORF030	1	25531..25995	154	ttttcaaggaggaggtcgctcgta	ctg	tga
100763	96ORF031	2	23633..24097	154	tttagtattgaaggtgattctgttag	atc	tga
100764	96ORF032	-2	2248..2706	152	ataagacacccaagggttttgccgc	atg	tga
100765	96ORF033	-3	39147..39605	152	agcatataaatcggtttagtcttgc	ttg	tga
100766	96ORF034	2	13161..13615	144	tagaagtgcgaagggtggaggcaat	ata	tga
100767	96ORF035	2	10628..11053	141	gagctaggattgcaggcaacgatat	ttg	tga
100768	96ORF036	2	24110..24535	141	gtatttttcataagggtgtttaa	atg	tga
100769	96ORF037	1	12583..12996	137	atgagggaacagagcaacaaatttt	att	tga
100770	96ORF038	1	15628..16032	134	atgttaagaatgatgcctagtgttaa	ttg	tga
100771	96ORF039	3	39816..40220	134	ctaatacactttacttaattaaagg	gtg	tga
100772	96ORF040	-3	27528..27932	134	tttccataaaataacgaggacacca	atg	tga
100773	96ORF041	3	16206..16607	133	gatgagggcgagggtgtcagagtag	atg	tga
100774	96ORF042	2	35720..36106	128	aagttactataactaaaattatggt	gtg	tga
100775	96ORF043	-2	35713..36081	122	ttaaacgtccccctcagttattgct	ttg	tga
100776	96ORF044	-3	9460..9828	122	agatccatcagttgagataaattc	ata	tga
100777	96ORF045	-3	5139..5504	121	ttctttttgtattctgtaatttca	att	tga
100778	96ORF046	2	11513..11872	119	aagttaattgtatagaggtggaaata	atg	tga
100779	96ORF047	2	22991..23350	119	gtcgtactacgtctgtataagagcga	gtg	tga
100780	96ORF048	3	8607..8963	118	tggaaaaaggaattgagtgatgacta	atg	tga
100781	96ORF049	1	23353..23697	114	atccggtttaaaccataaaggtagag	gtg	tga
100782	96ORF050	-2	2728..3072	114	tggtaaatattgattacattaaata	ata	tga
100783	96ORF051	3	4692..5021	109	tcaaaatatacggagggtagtcaact	atg	tga
100784	96ORF052	-1	20882..21211	109	qtgacaaagagacaactaaaaagtc	gtg	tga
100785	96ORF053	1	40252..40578	108	acgactaaatttttttagtcgttttt	att	tga
100786	96ORF054	1	4942..5262	106	aactataaaactaaaaacaaatttt	atg	tga
100787	96ORF055	-2	4840..5151	103	ccgtcgcaattatagttcgcttaa	atc	tga
100788	96ORF056	3	36324..36623	99	aatttaacacaaagtaggtggccta	atg	tga
100789	96ORF057	2	1394..1690	98	cttcagtcggctcttttttagcattaa	atc	tga
100790	96ORF058	-3	26247..26537	96	tacttctttttctataatccgacca	att	tga
100791	96ORF059	-1	21485..21772	95	agactcaacggctttttgaacatac	ttg	tga
100792	96ORF060	-3	22647..22931	94	cctctttgcaaccgacaagagctga	ata	tga
100793	96ORF061	1	14023..14304	93	ttatcttaattaaagggtgacagatga	gtg	tga
100794	96ORF062	-2	38281..38559	92	tataaacttagcgattgtacttgc	ttg	tga

200

100795	96ORF063	-3	30786..31064	92	gtctcctaactacatcttgcctta	gtg	tga
100796	96ORF064	-2	30205..30480	91	atgcattctacttttggatgtaatac	ata	tag
100797	96ORF065	-1	2617..2886	89	aaggtcttaataaaaaattctcccttc	ttg	tta
100798	96ORF066	3	28056..28125	89	aaggtcttagtcggctggttaactga	att	tta
100799	96ORF067	-3	17142..17411	89	ttccgtttattgcgtcgtaagttgt	ttg	tga
100800	96ORF068	2	12326..12589	87	aatgcacgtcgctttgctcgccctaa	ttg	tag
100801	96ORF069	2	42734..42997	87	tttttaggcacacgatataaatgaaa	gtg	tta
100802	96ORF070	1	11869..12129	86	aaatggtcagaataggagtgagc	ata	tta
100803	96ORF071	3	15396..15656	86	aacaaagctatacaaatatcgataa	att	tta
100804	96ORF072	-3	37749..38009	86	agattttttcgggttaccocctagac	att	tta
100805	96ORF073	3	11244..11501	85	acatgcataatagagggtggaataa	alg	tag
100806	96ORF074	-3	42936..43193	85	aattatttaacttactaattttcttc	ttg	tta
100807	96ORF075	-3	26610..26667	85	tactgccaatgtcccatcttcaacc	att	tta
100808	96ORF076	-1	11126..11380	84	tttatctaatacatttaagttcaacc	atc	tta
100809	96ORF077	-2	16537..16791	84	taccaccataataggcaggtagtag	gtg	tag
100810	96ORF078	-3	19521..19775	84	aataactttgaattgatacctcaac	ata	tga
100811	96ORF079	3	13608..13859	83	ttagggtcaaatggagggagacacaa	atg	tag
100812	96ORF080	-3	28029..28280	83	tgagaagtcgcccgaagcaactga	att	tga
100813	96ORF081	3	20973..21221	82	agattgaattatcccatctcatgactt	atc	tag
100814	96ORF082	-1	8729..8974	81	cgattattgtgttttcaatttcaaa	ttg	tga
100815	96ORF083	-3	3147..3392	81	tttagcctttatataatacaactcttc	gtg	tga
100816	96ORF084	3	1611..1853	80	tgttttatcttttagtttctttcttt	ttg	tga
100817	96ORF085	-2	29470..29709	79	ctcttatcaccttcgtttttaggca	atc	tta
100818	96ORF086	1	35188..35424	78	gcgcgaagtcgatttggatatttcaa	ctg	tag
100819	96ORF087	-2	13039..13275	78	ttttgatgagctcttaagtgcttt	ttg	tag
100820	96ORF088	3	24930..25163	77	gaactatcatataaagctaaatgga	ata	tga
100821	96ORF089	-3	23329..23562	77	tccagataaagatagtgtaattccc	ata	tta
100822	96ORF090	-3	16603..17036	77	accttttagtcgaataccctcgctca	ata	tag
100823	96ORF091	-1	22559..22789	76	aacgcttctggttttaacgttcatgt	atg	tta
100824	96ORF092	3	18360..18587	75	attgcraaagataattgtaagtagat	atg	tta
100825	96ORF093	-2	25384..25608	74	catgatttctcttgaattctcttctc	atc	tta
100826	96ORF094	1	10417..10638	73	aacacacataataggaggttataaaa	atg	tag
100827	96ORF095	3	12963..13184	73	tactaaacgaagataaaaactatgac	att	tta
100828	96ORF096	1	42994..43212	72	gatcgcttgaanaacgaagaagataa	ata	tta
100829	96ORF097	-1	36047..36265	72	tcaagcattacacacctgtgacttttc	atc	tta
100830	96ORF098	-2	36766..36984	72	caggttccggttcaaaatccagatga	ata	tta
100831	96ORF099	-2	34765..34983	72	tcaattctttttataaaaacgggtacc	atg	tag
100832	96ORF100	1	10198..10413	71	acaagaagactcagaggtttttcac	atg	tta
100833	96ORF101	1	15208..15423	71	gggaacaaagtttaagataaaggagag	atg	tga
100834	96ORF102	3	4209..4424	71	atttttaaaacgaatataggagagg	ctg	tag
100835	96ORF103	3	11673..11888	71	catgcaccttatgttatgcgtctag	ctg	tta
100836	96ORF104	3	12117..12332	71	tttagctcaaaagagcttttgactt	gtg	tta
100837	96ORF105	3	23892..24107	71	gatggctgggttatccaggtttataa	gtg	tta
100838	96ORF106	-3	34428..34643	71	tagacttttggcaatttgttgttga	att	tta
100839	96ORF107	-3	24495..24710	71	ggcacattaccaattgttaattttaa	atg	tta
100840	96ORF108	-1	23876..24088	70	acataatttaaccacctctatgaaaa	ata	tta
100841	96ORF109	-2	17317..17529	70	acctgtacgcttttgcctcggtatta	att	tta
100842	96ORF110	-3	38931..39143	70	accttcattcttttctgagtgaagaa	atg	tta
100843	96ORF111	-3	21855..22067	70	agtaaattttttctttctgtgtctc	att	tga
100844	96ORF112	1	3217..3426	69	aaatgtcaacgggaggtgatacga	atg	tta
100845	96ORF113	-1	25469..25678	69	tcagggtatattccctaatacttag	ctg	tta
100846	96ORF114	-2	9838..10047	69	ataataatcatcacggtaaaatgac	atc	tga
100847	96ORF115	1	13819..14022	67	gcagtaggggttatggcaggctaaag	ttg	tga
100848	96ORF116	-1	41033..41236	67	caacttcatgacctgcatgtctttaa	ata	tta
100849	96ORF117	-3	24711..24914	67	tctgtgtattccatttaactttta	atg	tta
100850	96ORF118	-1	12374..12574	66	tccatctccctcaaaataaagctgg	ttg	tta
100851	96ORF119	-1	3980..4180	66	ctcctatttttctgttttaaatcttc	att	tga
100852	96ORF120	-3	6033..6233	66	ttgtaatttaaaataataacgataa	ata	tta
100853	96ORF121	-2	37939..38136	65	ctgaaatgcttggatatttgccttaa	att	tga
100854	96ORF122	2	37892..38086	64	acgacaaaacacaaataaagatta	gtg	tga
100855	96ORF123	-3	29193..29387	64	ggacgtctgacttttaaatgtgaagc	ata	tga
100856	96ORF124	1	4408..4599	63	tttatcggtaccaatttaattgatta	atg	tta
100857	96ORF125	-1	7787..7978	63	ttaaaaatccaaagttttgcatctgt	att	tga
100858	96ORF126	-3	27027..27218	63	aaatttgaacaaaggcatatttga	gtg	tga
100859	96ORF127	3	15051..15239	62	atcgagtcagagaggttttggagaa	gtg	tga
100860	96ORF128	-1	6914..7102	62	agcgaaatgggtttgattgttgactc	ata	tga
100861	96ORF129	-3	31332..31520	62	tcttatttgcctctgcttctctataa	atg	tga
100862	96ORF130	-3	30084..30272	62	gaattcatcttcaccttcaactga	gtg	tta
100863	96ORF131	3	11058..11243	61	agaaaaagaaatgaagtgatcta	atg	tta
100864	96ORF132	-1	36434..36619	61	taagcatggttaactcactctttaa	ata	tga
100865	96ORF133	-1	35591..35776	61	ctaaactattgctgaacacgctcgt	att	tta
100866	96ORF134	-2	9250..9435	61	atccatgagcttataacccgctcta	att	tga

100867	96ORF135	1	29563..29745	60	cgacaactttttttaggactagtaa	gtg	tga
100868	96ORF136	-3	12486..12668	60	cactttacttttcaactctgttcagga	ttg	taa
100869	96ORF137	-1	14501..14680	59	caaaactgaagctaaagtaacagca	atc	tga
100870	96ORF138	-2	23326..23505	59	cttctgacattttagtgaatttttag	ttg	tga
100871	96ORF139	-3	42672..42851	59	aatccgggaatttttagcaattttat	atc	taa
100872	96ORF140	-3	31137..31316	59	acttgattgactagtadaagtcgtac	atg	taa
100873	96ORF141	-3	18969..19148	59	aacaaaaataacattataggatct	ata	taa
100874	96ORF142	-3	4740..4919	59	catataattttgttttttagttttat	att	tga
100875	96ORF143	2	36107..36283	58	acacaaatctgaggggagcgtttaa	atg	taa
100876	96ORF144	3	16029..16205	58	tatagcgaagctaaagaggtagataa	ata	tag
100877	96ORF145	-3	29013..29189	58	tgtcacgacgagatagctgcgaacc	att	tga
100878	96ORF146	-3	14883..15059	58	aactcttgaaatgtctgactaaagta	ttg	taa
100879	96ORF147	-1	18251..18434	57	catacagcgttaattgcaagtaattct	atg	taa
100880	96ORF148	-1	13583..13766	57	aataacctttcttaactgaatgttga	ata	taa
100881	96ORF149	-2	10756..10929	57	taaaattcacattctctactgatat	ctg	tgg
100882	96ORF150	2	14171..14341	56	atttttaataaggaaggtttattaa	ctg	tag
100883	96ORF151	2	19217..19387	56	cctacatctctattgagctactttt	atg	tga
100884	96ORF152	-1	12614..12784	56	atttctacagtaaaaatacttttat	ctg	taa
100885	96ORF153	-2	11836..12006	56	ttgcattacctatttgcgaatgctag	ttg	taa
100886	96ORF154	-2	4165..4335	56	atataacgcttttctgctcgaccaa	atc	tga
100887	96ORF155	-3	40464..40634	56	aaatcaggatcgaactgcttccctta	atg	tga
100888	96ORF156	3	423..590	55	tggttaattttgataatttttagctta	ata	taa
100889	96ORF157	-1	41879..42046	55	gtagcaaaattttttatttctaagtaa	ata	taa
100890	96ORF158	-2	36166..36333	55	cattcagtgctgctgctgtttggtaa	atc	tag
100891	96ORF159	-2	16228..16395	55	tttaacatctgagcacaactttttat	ttg	taa
100892	96ORF160	3	1038..1202	54	atctctaagcagttgttgagcagcg	ttg	taa
100893	96ORF161	-1	19193..19357	54	tccttctgtttaggtacaccaaaca	atg	tag
100894	96ORF162	-1	18074..18238	54	ctcgtctctattaacacaaatagatcc	ata	tga
100895	96ORF163	-1	15386..15550	54	agccatcacaggactgtaaaatttca	ctg	taa
100896	96ORF164	-1	10049..10213	54	tacatcgatttcaataagcttttga	att	tag
100897	96ORF165	-2	18514..18678	54	gtgcttcaatatcatcttatttactt	ata	taa
100898	96ORF166	-2	11104..11268	54	ctagccatgatttacccttaatttag	ttg	tag
100899	96ORF167	-3	13764..13928	54	agacagtttataatgtgtatctcta	ata	tga
100900	96ORF168	1	14305..14466	53	ttttgaaattttttaggagcagtaa	atg	tag
100901	96ORF169	-1	17885..18046	53	gtgttgagccttaataagactcttt	ata	tga
100902	96ORF170	-1	10790..10951	53	tagggcgtttacatataccagtttaa	att	taa
100903	96ORF171	-3	12765..12926	53	attctcgttttaagtataaaacgct	ctg	taa
100904	96ORF172	3	22836..22994	52	cgcttcgcaacgctttaaaccactga	ata	tga
100905	96ORF173	-1	15956..16114	52	ctctacatcatcattagccgtcgtc	ata	taa
100906	96ORF174	-1	10571..10729	52	tagtgccattcatattactttcttaa	ata	taa
100907	96ORF175	-1	3440..3598	52	cagccttatcttctacatcaaacatga	ttg	taa
100908	96ORF176	-3	37170..37328	52	tttatctaaaacgttgcgtgaagca	gtg	taa
100909	96ORF177	-3	6693..6851	52	ttcctaatctactaagtaactcgat	ata	taa
100910	96ORF178	-3	5655..5813	52	gacatcttgattagttttttcagtc	atc	tag
100911	96ORF179	1	34564..34719	51	gttacagctgaagctgataaaaatag	ttg	tag
100912	96ORF180	1	42661..42816	51	atataaattctaacacacaaaatact	atg	tga
100913	96ORF181	-2	37741..37896	51	tcgacgcactgtcaactgagttttt	atc	taa
100914	96ORF182	-2	25039..25194	51	ttcgttaacttttttctccgtcatta	att	tga
100915	96ORF183	-2	4534..4689	51	tcagtttttaatttttcagccatag	ttg	tga
100916	96ORF184	1	6721..6873	50	ggagctggagaactcacagtaaaag	ttg	tag
100917	96ORF185	2	36548..36700	50	acaaaaataacggatattgaaaat	gtg	taa
100918	96ORF186	-1	40025..40177	50	tggagatcctgataaataacatcatt	ata	tga
100919	96ORF187	-1	34466..34618	50	attaccttttaacaggctcagcgcca	ttg	tga
100920	96ORF188	-1	33842..33994	50	agttcctctatctgattcatagaaa	ctg	taa
100921	96ORF189	-1	24914..25066	50	acatagaatggtcttccggtgttga	atc	taa
100922	96ORF190	-2	20395..20547	50	tatcttagagataaccctcccactc	ata	tga
100923	96ORF191	3	24768..24917	49	aaaggaaattgaagcagtgaaacacg	ctg	taa
100924	96ORF192	-1	16169..16318	49	ttgtgggtttggcaacgttgcctgt	atg	tga
100925	96ORF193	-2	39100..39249	49	cagtaacggtttttaccggtgaccc	ttg	taa
100926	96ORF194	-2	25921..26070	49	ttggtacagacgcttcttgcataatc	ttg	taa
100927	96ORF195	-2	17779..17928	49	caaccaatgctcgggatggtcaggg	ttg	tga
100928	96ORF196	-2	14182..14331	49	tttaatactttttctctagcaatgc	atc	tga
100929	96ORF197	-2	7609..7758	49	ttatcatcaaacgacttaacaccaa	ttg	tga
100930	96ORF198	-2	1537..1686	49	ttattagctagtgctttagttag	gtg	taa
100931	96ORF199	-3	7719..7868	49	tastacttgratcggaatgctatct	att	taa
100932	96ORF200	2	22271..22417	48	ttcttttaatgaggttaaacctctaa	ttg	taa
100933	96ORF201	2	30353..30499	48	cttactattggcgaanaantaagcc	ttg	tag
100934	96ORF202	2	32591..32737	48	agattgaagcccaacggacattta	ttg	taa
100935	96ORF203	2	39131..39277	48	agcaaaagactttaaaagaaataag	ata	tag
100936	96ORF204	-2	36985..37131	48	atcttccctgagaaactgtcccaact	att	tga
100937	96ORF205	-3	38721..38867	48	aaggaaacccctttacacacatcgtc	ata	taa
100938	96ORF206	-3	35880..36026	48	gttaacatagcgttttctgctgca	att	taa

100939	96ORF207	-3	11550..11696	48	tcgctctctcgcctccatgattttgg	ata	taa
100940	96ORF208	2	37178..37321	47	cgatttgaagaccaccccttatgtaa	gtg	caa
100941	96ORF209	2	42341..42484	47	tcgctattttaaaccacccatactag	ttg	taa
100942	96ORF210	3	41850..41993	47	aaaggttaataacgtaaggagcggt	att	tag
100943	96ORF211	-1	6662..6805	47	tcgttggaaatggtgggacgaattgg	ttg	tga
100944	96ORF212	-2	25213..25356	47	agtagcacattcccaaaattgttaa	atc	taa
100945	96ORF213	-3	42219..42362	47	gtggcttgatcatttataataaac	ata	taa
100946	96ORF214	3	27834..27974	46	aaaagatttttagacttcgttagaac	atc	tag
100947	96ORF215	3	35811..35951	46	ttacgcaatagtttagatgtagacg	ata	taa
100948	96ORF216	-1	5402..5542	46	tttcggtaaggtgtattccaacttga	att	tga
100949	96ORF217	-2	24229..24369	46	tataggtctgttaagcacataaact	atc	taa
100950	96ORF218	-2	6253..6393	46	ttgtcattcttgcataacgctcaga	ttg	taa
100951	96ORF219	1	883..1020	45	aaatcactcccaaatattcgttaa	ata	taa
100952	96ORF220	2	32936..33073	45	gataaagggtatagacaaagtattgt	atc	taa
100953	96ORF221	3	41703..41840	45	ggttaagcctataggtggtttgttag	ctg	taa
100954	96ORF222	-1	39860..39997	45	acttttattaggttcaactccattt	att	taa
100955	96ORF223	-1	24716..24853	45	acatttcaaatgattctggacaac	ata	taa
100956	96ORF224	-2	26794..26931	45	caatatccagccatgtagtttttaa	ctg	taa
100957	96ORF225	-2	19201..19338	45	caaacaatggattgtatcaaataa	atg	tga
100958	96ORF226	-2	15709..15846	45	tgacttgcctgtgttcttaacacaa	ata	taa
100959	96ORF227	-3	36711..36846	45	acattgactgcccgcgaataattatct	ata	tga
100960	96ORF228	3	2325..2459	44	tcgccatagtgagttccaataccgt	ata	taa
100961	96ORF229	-1	38612..38746	44	ttgtcattgatacctattcttataag	atg	tga
100962	96ORF230	-1	31733..31867	44	gctggattgtatggcttaaaagtaat	ctg	tag
100963	96ORF231	-2	12076..12210	44	tgactcatagcttttaactgttcgt	ctg	taa
100964	96ORF232	-3	31644..31778	44	atagttcccaagtgtaaccctagt	ttg	taa
100965	96ORF233	-3	23988..24122	44	atttgatttgaagttcaggctcaa	ctg	taa
100966	96ORF234	-3	17529..17663	44	agtagcttttttgaatcgtacctta	atg	taa
100967	96ORF235	1	7153..7284	43	aatgcttaattggtccaatagaataca	atg	tag
100968	96ORF236	2	2681..2812	43	ttcttccacttcaacttcaacttcc	ata	tga
100969	96ORF237	2	4496..4627	43	gtactatggtttcacagttcttagcga	ttg	taa
100970	96ORF238	-1	41720..41851	43	caactgttaattcttgatatttagtga	ata	tga
100971	96ORF239	-1	35324..35455	43	acttactaataaataagataagttt	gtg	taa
100972	96ORF240	-1	8570..8701	43	atccccctgtttgacttaatacatca	atc	tga
100973	96ORF241	-2	33502..33633	43	ataattttgtaataactcttagggat	atg	tag
100974	96ORF242	-2	23662..23793	43	agctaatgctacagcagttgtgttaa	atc	tag
100975	96ORF243	-3	32391..32522	43	acctggacgagcttgcgtcatataa	ata	cag
100976	96ORF244	-3	30273..30404	43	aaaactttcgttataactcttggttaa	atc	tga
100977	96ORF245	-3	5895..6026	43	tcgactaataatgcttataattctta	atc	taa
100978	96ORF246	-3	2679..2810	43	attcatcaagaataatagccggtc	atg	tga
100979	96ORF247	1	34891..35019	42	acatcaagcaaatctggtgtgttag	ttg	taa
100980	96ORF248	2	30668..30796	42	aattattacattaaagctgggtgga	atg	tag
100981	96ORF249	2	31838..31966	42	caaatattagcttctgagtgttag	atg	taa
100982	96ORF250	2	33519..33647	42	cttaccgaanacacacaggtagaa	ata	taa
100983	96ORF251	-1	20486..20614	42	cttccgtacagacccacacgaatga	ttg	tga
100984	96ORF252	-1	15128..15256	42	gataattccattactagctactacta	ata	tga
100985	96ORF253	-2	41446..41574	42	aaaaccttaattcagatataacgataa	ttg	tga
100986	96ORF254	-2	41005..41133	42	gttataaacatgaccggtacaaagc	ata	taa
100987	96ORF255	-2	23008..23136	42	aggataaatgacttgacctcttctc	ata	taa
100988	96ORF256	-2	14794..14922	42	ttgtatgctcagttagttgttga	ttg	tag
100989	96ORF257	-2	8503..8631	42	tacctaaactttttaaataattctta	atg	tga
100990	96ORF258	-3	22143..22271	42	aaacgctttgttaaaatgcctctgca	att	tga
100991	96ORF259	-3	18639..18767	42	cttgatctatttatagagattaac	att	tag
100992	96ORF260	-3	15624..15752	42	gttttggtaactagccactgtatag	ata	taa
100993	96ORF261	2	18746..18871	41	catattcaggtcttaataaggtcac	ata	taa
100994	96ORF262	-1	13067..13192	41	aatttaatttaattcttcttctgttg	ttg	taa
100995	96ORF263	-2	18742..18867	41	taacagacagctcctaactgccttac	att	tga
100996	96ORF264	-2	18376..18501	41	catattatcataaagaaacaaagtaac	ttg	taa
100997	96ORF265	-2	367..492	41	ctaaacgaaagaaggggtacataac	atc	tga
100998	96ORF266	-3	32802..32927	41	aggtatattccatttgatacaact	ttg	taa
100999	96ORF267	-3	10194..10319	41	atcatcgaaaggcgaataactcgtta	ttg	tga
101000	96ORF268	1	1159..1281	40	ttattcttcttcttctgaattgttaa	atg	taa
101001	96ORF269	2	10373..10495	40	gacagaggtgaaagaaatcatga	atg	taa
101002	96ORF270	2	15734..15856	40	ttattcggcgaatactgcactgtatg	ttg	tag
101003	96ORF271	-1	43451..43573	40	c c tNo shine-dalgarno sequence	att	tga
101004	96ORF272	-1	36959..37081	40	acgtatcaaaataaacttttattag	atg	tag
101005	96ORF273	-1	35798..35920	40	ctgacgactcttctgtgtttgtagc	att	taa
101006	96ORF274	-1	8147..8269	40	ctgtctctctatgtttgttagtct	ctg	tga
101007	96ORF275	-2	43066..43188	40	tttaacttactaattttcttttgat	ata	tga
101008	96ORF276	-2	42535..42657	40	aaataatgcaaatgtttcttcatagt	att	tag
101009	96ORF277	-2	30628..30750	40	tttgtagtccgctctctgcaaaagt	ctg	taa

203

101010	96ORF278	-2	13391..13413	40	ttcgtatcttccaagcaattcattt	ttg	tga
101011	96ORF279	-2	3172..3294	40	cagattgttttagtaacgcctaat	atc	taa
101012	96ORF280	-3	18804..18926	40	taataaaccaacacgtgtatcaaca	att	tag
101013	96ORF281	-3	15843..15965	40	atttataaaagtgattctcataacca	atc	tag
101014	96ORF282	-3	8460..8582	40	ttagtcatcactcaattctttttcc	att	tga
101015	96ORF283	-3	7593..7715	40	gatgttctacacagtgcttaaacac	atg	tga
101016	96ORF284	-3	6453..6575	40	aattaatttttaattaccatttcta	att	tga
101017	96ORF285	1	15082..15201	39	caatacttgaacacacatttcaaaag	att	tga
101018	96ORF286	1	34444..34563	39	acacaaacgttaataagcaaaagtg	atg	tag
101019	96ORF287	2	27920..28039	39	cctatttttagcagttgttgcagtda	ttg	tag
101020	96ORF288	2	28415..28534	39	atcggtcttttaactggcgtaaacg	atc	tag
101021	96ORF289	2	38147..38266	39	tatcaaatgcttaatttaggaagc	atc	tga
101022	96ORF290	3	40917..41036	39	gcataatttaaacacttccacatcat	atg	tga
101023	96ORF291	-2	38815..38934	39	tcctataaaacagcgttacacgcgaac	ata	tga
101024	96ORF292	-2	32671..32790	39	ctataggattataaatcgctgacgt	ata	tga
101025	96ORF293	-2	31216..31235	39	ttgatttgatgtttcttataacttga	ttg	tga
101026	96ORF294	-2	21589..21708	39	gtatcttcatcagaatcgcttaaaa	atc	tga
101027	96ORF295	-2	18976..19095	39	tatcaatatatgctaacctagcacc	ata	tga
101028	96ORF296	-2	11482..11601	39	gccactctgactctttttgcaacc	att	tga
101029	96ORF297	-3	12933..13052	39	tcaagaaataatgtttcttcaattt	ata	tga
101030	96ORF298	-3	8262..8381	39	gaactgactctgcttaaatgattta	att	tag
101031	96ORF299	-3	6993..7112	39	cattagcatttagcgaatgggtttga	ttg	tga
101032	96ORF300	2	23516..23632	38	actacatctgaacaaactaaaatttc	atc	tag
101033	96ORF301	2	25943..26059	38	agattagagagaaagaaagagagac	gtg	tga
101034	96ORF302	2	36929..37045	38	tattgggttttttgaacatggggca	atg	tag
101035	96ORF303	3	4476..4592	38	atbaaagctaccttagtagcagtaact	atg	tga
101036	96ORF304	3	20586..20702	38	tactctaaagtagctaaagcaatcac	gtg	tga
101037	96ORF305	3	28356..28472	38	cggttaccatgtgcttgcagcat	ttg	tga
101038	96ORF306	-1	24359..24475	38	acttaataaaagcgtatcgtgcc	atg	tga
101039	96ORF307	-1	20147..20263	38	ttgactctatacagatttaactcctt	att	tag
101040	96ORF308	-2	38156..38274	38	ttcgtatccactttcttaagaagac	gtg	tga
101041	96ORF309	-2	35149..35265	38	agctgtgttgcattcgttttaacga	ata	tga
101042	96ORF310	-2	31423..31539	38	gtaatatgattaggtctcctcttat	ttg	tga
101043	96ORF311	-2	10438..10554	38	cgcttttaaatcgttttaggtcact	atc	tga
101044	96ORF312	-2	1390..1506	38	gagaacaacacaaacatttaacaca	atc	tga
101045	96ORF313	-3	33051..33167	38	acgtctctgtttcttagatcgttaaac	ata	tag
101046	96ORF314	-3	25194..25310	38	agcaaacogtttaagataaacattga	atc	tga
101047	96ORF315	-3	6273..6389	38	cattcttctgaacacgtcagatttga	ctg	tga
101048	96ORF316	-3	4281..4397	38	ataattcgtattcatttaactttaa	att	tag
101049	96ORF317	1	2260..2373	37	atgactcctttttctcatatttcttt	ata	tga
101050	96ORF318	2	21230..21343	37	arttcacacttttttagtgcctctt	ttg	tga
101051	96ORF319	3	18018..18131	37	atctagagtcacaaatttaaacctcg	atg	tag
101052	96ORF320	3	36972..37085	37	attacagacatccttaaggtttcccg	att	tga
101053	96ORF321	-1	36302..36415	37	ctcttgagtttttttgacctaaactca	atc	tga
101054	96ORF322	-1	32606..32719	37	ccataagttatttctccagttccat	att	tga
101055	96ORF323	-1	11453..11566	37	ttaaaacgttctttttttatcaattc	att	tga
101056	96ORF324	-1	7268..7381	37	tactgggttgcacccagtgaaattct	ata	tga
101057	96ORF325	-2	32347..32460	37	ttactgcatttgcattatggcgataa	atc	tag
101058	96ORF326	-2	24682..24795	37	acgtttattagctcctataaagccat	ata	tag
101059	96ORF327	-2	23905..24018	37	aaatggctgtggcgcttgaccatatt	gtg	tga
101060	96ORF328	-2	21460..21573	37	agagcacttaatacgtttttgtcttt	ctg	tga
101061	96ORF329	-2	21208..21321	37	gaacttaacttcttcgattctcatat	atc	tga
101062	96ORF330	-2	18085..18198	37	ccagtcgaacacagcaaaagtattct	ttg	tag
101063	96ORF331	-2	8170..8283	37	actttgagacgtcgtctgtctctct	atg	tag
101064	96ORF332	-2	5971..6084	37	caatttcttttctcgttttctcttag	ttg	tag
101065	96ORF333	-3	37632..37745	37	accttgcttaatacagtcgttaacta	att	tga
101066	96ORF334	-3	29628..29741	37	ctgagttagtggttgaataatgctcat	ttg	tag
101067	96ORF335	-3	7164..7277	37	ttagcggatctcgttttcttagtaa	atc	tga
101068	96ORF336	1	23903..23913	36	gtaaaaaagacaatatgactatta	ctg	tga
101069	96ORF337	1	43258..43368	36	taattgacggtgtatttttttaggt	ttg	tga
101070	96ORF338	2	12668..12778	36	gaactggtggaatggcatggaaaca	atc	tag
101071	96ORF339	2	28292..28402	36	ttcactgcttttaattcagttgctta	ctg	tga
101072	96ORF340	2	35396..35506	36	ttcctaatgaacataaagtcacaggt	att	tga
101073	96ORF341	3	25428..25538	36	actcagagacaattagaanaagcaa	ttg	tga
101074	96ORF342	-1	40913..41023	36	tatctgggaattttaatctaatcaaa	ata	tga
101075	96ORF343	-1	39173..39283	36	tgcacatttttagtgctcaggattga	ttg	tga
101076	96ORF344	-1	37580..37690	36	gggtctaccttttaacgtcgtttcag	ata	tga
101077	96ORF345	-1	31556..31666	36	ggattattcttttctaataacttcaa	ttg	tga
101078	96ORF346	-1	29972..30082	36	ggctactccttatctaaaataataat	ttg	tga
101079	96ORF347	-1	28787..28897	36	ctgcacaaagtcgttagcaattactt	ttg	tga
101080	96ORF348	-1	21839..21949	36	ttaaaactccgataaaaataacattgc	ctg	tga
101081	96ORF349	-1	3647..3757	36	taaaaactccgaattaccacagcgt	ttg	tga

55

204

101082	96ORF350	-2	40801..40911	36	accatttccaattttgcctatattat	gtg	tag
101083	96ORF351	-2	38953..39063	36	tatctttttaaatttccgtatagc	atc	taa
101084	96ORF352	-2	31585..31695	36	tagctgtcatcactagtatttttga	atc	taa
101085	96ORF353	-2	24550..24660	36	atagtcggtttttacggcctcgtact	att	tag
101086	96ORF354	-2	20083..20193	36	atcatcatttttgatattttctaaac	ata	tga
101087	96ORF355	-2	991..1101	36	gcattctggcagtagacgtaaaac	atc	tga
101088	96ORF356	-3	38148..38258	36	taagaaagcgtgcgcgagtcataa	att	tga
101089	96ORF357	-3	8790..8900	36	tcgaagttatcttagcgtatctttct	ttg	tag
101090	96ORF358	-3	4458..4568	36	ctcataaaagtattctttctagtag	atg	tag
101091	96ORF359	1	4666..4773	35	ttatcacaataatcacacttaataa	atc	tag
101092	96ORF360	1	11569..11676	35	ataaattacccgaacatgaaatga	att	tga
101093	96ORF361	2	6132..6239	35	ggaaaaacaaatgattgttctagtag	ttg	tga
101094	96ORF362	-1	40418..40525	35	ttctgtaggtcctactactcttttaa	ttg	tag
101095	96ORF363	-1	34358..34465	35	gttttctgtgattctgattctgtga	atg	tga
101096	96ORF364	-1	20654..20761	35	ctatttccactgattccccatctaa	atg	tga
101097	96ORF365	-1	8423..8530	35	ctctttttgagattacgaggtttca	att	tag
101098	96ORF366	-1	2402..2509	35	tgacgtatcggaacatttttagatca	atc	taa
101099	96ORF367	-2	36607..36714	35	aaaataaaaaagccagtgccgaagca	ctg	tag
101100	96ORF368	-2	27061..27168	35	caaatcgtctcgcagcgttcaataa	atc	tag
101101	96ORF369	-2	26470..26577	35	atgagttgttaagtttaccccaaat	atc	taa
101102	96ORF370	-2	10227..10434	35	ccgtgcacattcttctcgggtataagta	ata	taa
101103	96ORF371	-2	8650..8757	35	gggtacgggttggttactggtgatat	atc	taa
101104	96ORF372	-3	14382..14489	35	gttcttttaattgattctactgttaa	att	taa
101105	96ORF373	-3	8151..8258	35	atgtttgttattctctgtgttagtct	atg	tga
101106	96ORF374	-3	5007..5114	35	aaacgattcgaagtggaacattatcc	ata	taa
101107	96ORF375	2	30563..30667	34	cgattagaattctttaaanaaggac	ttg	tga
101108	96ORF376	-1	13916..14020	34	tctatgtcaggtcaatttgcactaa	att	taa
101109	96ORF377	-1	9236..9340	34	cttttctgttgaatattgtttttaa	atc	taa
101110	96ORF378	-1	9026..9130	34	actctttatcttttagttgttttaa	ata	tag
101111	96ORF379	-2	28447..28551	34	cttttctgataataaagtttagtgc	ttg	tga
101112	96ORF380	-3	40329..40433	34	ccatttaccctttctgagattgttga	ttg	tga
101113	96ORF381	-3	39801..39905	34	caaaaqctgaaggctttttccatcc	ttg	taa
101114	96ORF382	-3	33831..33935	34	atgttgtttgttaactcgtattaagtt	atc	tga
101115	96ORF383	-3	33687..33791	34	gttattacgtcttaataacttctgttt	gtg	tag
101116	96ORF384	-3	13530..13634	34	tatacgcactagtagcactga	ttg	tga
101117	96ORF385	-3	3843..3947	34	tttgattgattgttctagtagaa	att	taa
101118	96ORF386	1	12256..12357	33	agtcataaagaagtagcaatgtga	ttg	tag
101119	96ORF387	2	2207..2308	33	tccaaagactcttttaactgttaact	atc	tag
101120	96ORF388	2	2519..2620	33	attgttgatttcgactgacttaaa	atg	tga
101121	96ORF389	2	22517..22618	33	agaagtaaatgcgttaattgtttag	atg	tag
101122	96ORF390	2	27302..27403	33	ttccaaaatttggtcaatagtgtag	ctg	taa
101123	96ORF391	2	32384..32485	33	actaaaaaggttgagaagcctgtag	atg	taa
101124	96ORF392	2	39287..39388	33	aaaaacggtactgtatcatcaatca	atc	tag
101125	96ORF393	3	18153..18254	33	gtagcatatgccgactttgatttga	atg	taa
101126	96ORF394	3	24189..24290	33	tcagacccttaacatttaacaaactag	ttg	tga
101127	96ORF395	-1	15266..15367	33	tcgataatttgaatagctgtttta	atg	cag
101128	96ORF396	-2	32339..32340	33	cttttagtgaaagcatcttagttga	ata	cag
101129	96ORF397	-2	16123..16224	33	ttatgtgtgcctatcatatttaacaa	ttg	tag
101130	96ORF398	-2	13648..13749	33	lcttttaactgaattgttaataagcat	ttg	tag
101131	96ORF399	-2	10987..11088	33	acttctgtaggattctttatatcaa	ttg	tga
101132	96ORF400	-2	3382..3483	33	cttactggttaattcttcaaaattaa	atg	taa
101133	96ORF401	-3	40794..40895	33	ccatatgatgtgaaagtggttaaat	ttg	taa
101134	96ORF402	-3	39978..40079	33	atattcctaaatcaactgaacctaa	att	tga
101135	96ORF403	-3	38607..38708	33	atcttcagtgtaaatcgacagcca	atg	tag
101136	96ORF404	-3	21288..21389	33	cagacacggtcttaagtccttttag	ata	taa

Table 11

SEQUENCE INFORMATION FOR PHAGES MATCHING WITH TABLE 1

M32695

Bacteriophage PM2 nuclease cleavage site

gi166145|gb|M32695|BM2NCS [166145]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M32693

Bacteriophage PM2 Hind III fragment 4

gi166144|gb|M32693|BM24HIND3 [166144]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M32693

Bacteriophage PM2 Hind III fragment 4

gi166144|gb|M32693|BM24HIND3 [166144]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M32694

Bacteriophage PM2 Hind III fragment 3

gi166143|gb|M32694|BM23HIND3 [166143]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

M26134

Bacteriophage PM2 structural protein gene containing purine/pyrimidine rich regions and anti-Z-DNA-IgG binding regions, complete cds

gi289360|gb|M26134|BM2PROTIV [289360]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

J02452

bacteriophage fi 3'-terminal region rna

gi215409|gb|J02452|PFITR3 [215409]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

AF020798

Bacteriophage Chp1 genome DNA, complete sequence

gi217761|dbj|D00624|BCP1 [217761]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 12 protein links, or 1 genome link)

X72793

Clostridium botulinum C phage BONT/C1, ANTP-139, ANTP-33, ANTP-17, ANTP-70

genes and ORF-22

gi516171|emb|X72793|CBCBONT [516171]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 6 protein links, or 4 nucleotide neighbors)

X51464

Clostridium botulinum D Phage C3 gene for exoenzyme C3

gi14907|emb|X51464|CBDPE3 [14907]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

D90210

Bacteriophage c-st (from C. botulinum) C1-tox gene for botulinum C1 neurotoxin

gi217780|dbj|D90210|CSTCITOX [217780]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

S49407

type D neurotoxin [bacteriophage d-16 phi, host = C. botulinum, type D, CB16, Genomic, 4087 nt]
gi260238|gb|S49407|S49407 [260238]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

X53370

Bacteriophage phi29 temperature sensitive mutant TS2(98) DNA polymerase gene
gi15733|emb|X53370|POTS298 [15733]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 7 nucleotide neighbors)

X53371

Bacteriophage phi29 temperature sensitive mutant TS2(24) DNA polymerase gene
gi15733|emb|X53371|POTS224 [15733]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 7 nucleotide neighbors)

X05973

Bacteriophage phi29 prohead RNA
gi15680|emb|X05973|POP29PRO [15680]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, or 4 nucleotide neighbors)

V01155

Left end of bacteriophage phi-29 coding for 15 potential proteins Among
these are the terminal protein and the proteins encoded by the genes 1, 2 (sus), 3, and (probably) 4
gi15659|emb|V01155|POP29B [15659]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 16 protein links, or 16 nucleotide neighbors)

X73097

Bacteriophage phi-29 left origin of replication
gi312194|emb|X73097|BP29ORIL [312194]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 5 nucleotide neighbors)

M14430

Bacteriophage phi-29 gene-17 gene, complete cds
gi215321|gb|M14430|P29G17A [215321]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 6 protein links, or 8 nucleotide neighbors)

M14431

Bacteriophage phi-29 gene-16 gene, complete cds
gi215319|gb|M14431|P29G16A [215319]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 7 nucleotide neighbors)

M20693

Bacteriophage phi-29 DNA, 3' end
gi215343|gb|M20693|P29REPINB [215343]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 nucleotide neighbors)

M21016

Bacteriophage phi-29 DNA, 5' end
gi215342|gb|M21016|P29REPINA [215342]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

5

207

M12456

Bacteriophage phi-29 genes 9, 10 and 11 encoding p9 tail, incomplete, p10 connector, complete, and p11 lower collar, incomplete, respectively
gi|215338|gb|M12456|P29P9 [215338]

10

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link,3 protein links, or 2 nucleotide neighbors)

M14782

Bacillus phage phi-29 head morphogenesis, major head protein, head fiber protein, tail protein, upper collar protein, lower collar protein, pre-neck appendage protein, morphogenesis(13), lysis, morphogenesis(15), encapsidation genes, complete cds
gi|215323|gb|M14782|P29LATE2 [215323]

15

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link,11 protein links, or 11 nucleotide neighbors)

M26968

Bacteriophage phi-29 (from Bacillus subtilis) proteins p1 delta-1 genes, complete cds, and the sus1(629) mutation
gi|341558|gb|M26968|P29P1D1A [341558]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link,2 protein links, or 1 nucleotide neighbor)

20

J02448

Bacteriophage f1, complete genome
gi|166201|gb|J02448|F1CCG [166201]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link,10 protein links,205 nucleotide neighbors, or 1 genome link)

25

M24832

Bacteriophage f2 coat protein gene, partial cds
gi|166228|gb|M24832|F2CRNACA [166228]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link,1 protein link, or 4 nucleotide neighbors)

30

J02451

Bacteriophage fd, strain 478, complete genome
gi|215394|gb|J02451|PFDCG [215394]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,5 MEDLINE links,10 protein links,204 nucleotide neighbors, or 1 genome link)

35

M34834

Bacteriophage f1 replicase gene, 5' end
gi|166139|gb|M34834|BFRREGRA [166139]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 9 nucleotide neighbors)

40

M38325

Bacteriophage f1 replicase gene, 5' end
gi|166137|gb|M38325|BFRREGRA [166137]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 9 nucleotide neighbors)

45

M35063

Bacteriophage f1 coat protein replicase cistron (R region) RNA
gi|166134|gb|M35063|BFRRCRRA [166134]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 3 nucleotide neighbors)

50

S66567

alpha-atrial natriuretic factor/coat protein=fusion polypeptide (human, bacteriophage f1, expression vector pFAN15, PlasmidSyntheticRecombinant, 510 nt)
gi|435742|gb|S66567|S66567 [435742]

55

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link,1 protein link, or 15 nucleotide neighbors)

50

55

X15031

Bacteriophage fr RNA genome

gi|15071|emb|X15031|LEBFRX [15071]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, 9 nucleotide neighbors, or 1 genome link)

U51233

Mus musculus neutralizing anti-RNA-bacteriophage fr immunoglobulin variable region light chain (IgM) mRNA, partial cds

gi|1277148|gb|U51233|MMU51233 [1277148]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 1669 nucleotide neighbors)

U51232

Mus musculus neutralizing anti-RNA-bacteriophage fr immunoglobulin variable region heavy chain (IgM) mRNA, partial cds

gi|1277148|gb|U51232|MMU51232 [1277148]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 1073 nucleotide neighbors)

U02303

Bacteriophage If1, complete genome

gi|3676280|gb|U02303|B2U02303 [3676280]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,10 protein links, or 1 genome link)

V00604

Phage M13 genome

gi|14959|emb|V00604|INM13X [14959]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 10 protein links, or 205 nucleotide neighbors)

A32252

Synthetic bacteriophage M13 protein III probe

gi|1567340|emb|A32252|A32252 [1567340]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

A32251

Synthetic bacteriophage M13 protein III probe

gi|1567339|emb|A32251|A32251 [1567339]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M12465

Bacteriophage M13 mp10 mutations in lac operon

gi|215210|gb|M12465|M13LACMUT [215210]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 215 nucleotide neighbors)

M24177

Synthetic Bacteriophage M13 (clone M13.SV.B12) SV40 early promoter region DNA

gi|209416|gb|M24177|SYNSVB12 [209416]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 nucleotide neighbor)

M24176

Synthetic Bacteriophage M13 (clone M13.SV.B11) SV40 early promoter region DNA

gi|209415|gb|M24176|SYNSVB11 [209415]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 nucleotide neighbor)

M24175

Synthetic Bacteriophage M13 (clone M13.SV.8) SV40 early promoter region DNA
gi208806|gb|M24175|SYNM13SV8 [208806]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 242 nucleotide neighbors)

M19979

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33
gi207813|gb|M19979|SYN33M13M [207813]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 617 nucleotide neighbors)

M19565

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33
gi207808|gb|M19565|SYN33M13H [207808]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 567 nucleotide neighbors)

M19564

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33
gi207807|gb|M19564|SYN33M13G [207807]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 12 nucleotide neighbors)

M19563

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33
gi207806|gb|M19563|SYN33M13F [207806]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 262 nucleotide neighbors)

M19561

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33
gi207804|gb|M19561|SYN33M13D [207804]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 27 nucleotide neighbors)

M19560

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33
gi207803|gb|M19560|SYN33M13C [207803]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

M19559

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33
gi207802|gb|M19559|SYN33M13B [207802]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 227 nucleotide neighbors)

M10568

Bacteriophage M13 replicative form II, replication origin, specific nick location
gi215220|gb|M10568|M13ORIB [215220]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 650 nucleotide neighbors)

M10910

Bacteriophage M13 gene II regulatory region and M13sj1 mutant
gi215209|gb|M10910|M13IIREG [215209]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 72 nucleotide neighbors)

M38295

Bacteriophage M13 HaeIII restriction fragment DNA
gi215208|gb|M38295|M13HAEIII [215208]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 67 nucleotide neighbors)

5

E02067 -210
DNA encoding a part of Bacteriophage M13 tg 127
gi2170311|dbj|E02067|E02067 [2170311]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

10

J02467
Bacteriophage MS2, complete genome
gi215232|gb|J02467|MS2CG [215232]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,8 MEDLINE links, 4 protein links, 20 nucleotide neighbors, or 1 genome link)

15

AJ004950
Bacteriophage P1 ban gene
gi3688226|emb|AJ011592|BP1011592 [3688226]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 protein link)

20

U88974
Bacteriophage P1 structural lytic transglycosylase (orf47), pep44b (orf44b),
pep44a (orf44a), and pep43 (orf43) genes, complete cds; and pep42 (orf42) gene, partial cds
gi2661099|gb|AF035607|AF035607 [2661099]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,5 protein links, or 1 nucleotide neighbor)

25

AJ000741
Bacteriophage P1 darA operon
gi2462938|emb|AJ000741|BPAJ7641 [2462938]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 10 protein links, or 31 nucleotide neighbors)

30

X01828
Bacteriophage P1 recombinase gene cin
gi15133|emb|X01828|MYP1CIN [15133]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

35

X98146
Bacteriophage P1 DNA sequence around the Op88 operator
gi1359513|emb|X98146|BP1OP88OP [1359513]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 nucleotide neighbor)

S61175
immI operon: icd=cell division repressor, antI=antirepressor (promoters
P51a, P51b) [bacteriophage P1, Genomic, 728 nt]
gi385908|gb|S61175|S61175 [385908]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 3 nucleotide neighbors)

40

X87824
Bacteriophage P1 gene 26
gi861164|emb|X87824|XCBP1G26 [861164]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 protein link)

45

X15638
Phage P1 DNA for lytic replicon containing promoter P53 and two open reading frames
gi15735|emb|X15638|PP1LREP [15735]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 24 nucleotide neighbors)

50

55

5

211

X17512

Bacteriophage P1 DNA for immunity region imm1
gi115479|emb|X17512|P1IMMUNITY [15479]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, or 4 nucleotide neighbors)

10

X16005

Bacteriophage P1 c1 gene for P1c1 repressor protein
gi115477|emb|X16005|P1C1 [15477]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

X03453

Bacteriophage P1 cre gene for recombinase protein
gi115135|emb|X03453|MYP1CRE [15135]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 12 nucleotide neighbors)

15

X06561

Bacteriophage P1 c1 gene 5'-region
gi115128|emb|X06561|MYP1C1 [15128]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 4 protein links, or 6 nucleotide neighbors)

20

V01534

Bacteriophage P1 genome fragment (IS2 insertion spot). This region contains
four unidentified reading frames and is known as insertion hot spot for IS2 insertion sequences
gi115118|emb|V01534|MYOVPI [15118]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 4 protein links, or 3 nucleotide neighbors)

25

X56951

Bacteriophage P1 gene10
gi406728|emb|X56951|BPP1GP10 [406728]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 3 protein links, or 1 nucleotide neighbor)

K02380

Bacteriophage P1 replication region including repA, parA, and parB genes and
incA, incB, and incC incompatibility determinants
gi215652|gb|K02380|P1IREP [215652]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 5 MEDLINE links, 4 protein links, or 8 nucleotide neighbors)

30

X87674

Bacteriophage P1 lydA & lydB genes
gi974763|emb|X87674|BACP1LYD [974763]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

35

X87673

Bacteriophage P1 gene 17
gi974761|emb|X87673|BACP117 [974761]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

40

M16618

Bacteriophage P1 c1 repressor binding sites
gi215600|gb|M16618|PP1C1 [215600]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

45

50

55

5

212

SEQ_PP1CTN

Bacteriophage P1 cin gene encoding recombinase, cixL recombination site, and 5' end of C invertible element
gi215607|gb|SEG_PP1CTN [215607]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

10

K03173

Bacteriophage P1 C invertible element, right end, and cixR recombination site
gi215605|gb|K03173|PP1CTN2 [215605]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

215605

15

Bacteriophage P1 cin gene encoding recombinase, cixL recombination site, and 5' end of C invertible element
gi215605|lcl|X01828 [215605]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

M25470

20

Bacteriophage P1 tail fiber protein gene, complete cds
gi341349|gb|M25470|PP1TFPR [341349]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 3 nucleotide neighbors)

M34382

Bacteriophage P1 sim region proteins, complete cds
gi215661|gb|M34382|PP1SIM [215661]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

25

M81956

Bacteriophage P1 R protein (R) gene, complete cds
gi215658|gb|M81956|PP1RP [215658]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 4 nucleotide neighbors)

M37080

30

Bacteriophage P1 mini-P1 plasmid origin of replication
gi215657|gb|M37080|PP1REPOR [215657]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 46 nucleotide neighbors)

M27041

Bacteriophage P1 ref gene, complete cds
gi215650|gb|M27041|PP1REF [215650]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

35

L01408

Bacteriophage P1 partition protein (parB) gene, 3' end
gi215642|gb|L01408|PP1PARB [215642]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 41 nucleotide neighbors)

SEG_PP1PAR

40

Bacteriophage miniplasmid P1 parA gene, 5' end
gi215639|gb|SEG_PP1PAR [215639]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 48 nucleotide neighbors)

M36425

45

Bacteriophage miniplasmid P1 parB gene, 3' end
gi215638|gb|M36425|PP1PAR2 [215638]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

50

55

5

M36424 213
Bacteriophage miniplasmid P1 parA gene, 5' end
gi215637|gb|M36424|PP1PAR1 [215637]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

10

M11129
Bacteriophage P1 miniplasmid origin of replication region
gi215632|gb|M11129|PP1ORDM [215632]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 43 nucleotide neighbors)

15

M25414
Bacteriophage P1 c1 repressor binding site, operator 88 (Op88)
gi215631|gb|M25414|PP1OP88A [215631]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 nucleotide neighbors)

20

M25413
Bacteriophage P1 c1 repressor binding site, operator 68 (Op68)
gi215630|gb|M25413|PP1OP68A [215630]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

25

M25412
Bacteriophage P1 c1 repressor binding site, operator 21 (Op21)
gi215629|gb|M25412|PP1OP21A [215629]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

30

M10510
Bacteriophage P1 recombination site loxR
gi215628|gb|M10510|PP1LOXR [215628]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

35

M10287
Bacteriophage P1 loxP X loxP recombination site
gi215627|gb|M10287|PP1LOXPX [215627]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

40

M10494
Bacteriophage P1 recombination site loxP
gi215626|gb|M10494|PP1LOXP [215626]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 134 nucleotide neighbors)

45

M10511
Bacteriophage P1 recombination site loxL
gi215625|gb|M10511|PP1LOXL [215625]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

50

M10512
Bacteriophage P1 recombination site loxB
gi215624|gb|M10512|PP1LOXB [215624]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

55

M10145
Bacteriophage P1 genome fragment with recombination site loxP
gi215623|gb|M10145|PP1CREX [215623]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 21 nucleotide neighbors)

5

214

M13327

Bacteriophage P1 Cln recombinase activated cross over site, junction IV, clone pSHI326
gi215622|gb|M13327|PP1CN26IV [215622]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

10

M13325

Bacteriophage P1 Cln recombinase activated cross over site, junction II, clone pSHI326
gi215621|gb|M13325|PP1CN26II [215621]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1401 nucleotide neighbors)

15

M13323

Bacteriophage P1 Cln recombinase activated cross over site, junction IV, clone pSHI325
gi215620|gb|M13323|PP1CN25IV [215620]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

20

M13321

Bacteriophage P1 Cln recombinase activated cross over site, junction II, clone pSHI325
gi215619|gb|M13321|PP1CN25II [215619]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1058 nucleotide neighbors)

M13324

Bacteriophage P1 Cln recombinase activated cross over site, junction I, clone pSHI326
gi215618|gb|M13324|PP1CN26I [215618]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

25

M13319

Bacteriophage P1 Cln recombinase activated cross over site, right junction, clone pSHI327
gi215617|gb|M13319|PP1CN27R [215617]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

30

M13320

Bacteriophage P1 Cln recombinase activated cross over site, junction I, clone pSHI325
gi215616|gb|M13320|PP1CN25I [215616]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

35

M13318

Bacteriophage P1 Cln recombinase activated cross over site, left junction, clone pSHI324
gi215615|gb|M13318|PP1CN24L [215615]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1370 nucleotide neighbors)

M13317

Bacteriophage P1 Cln recombinase activated cross over site, right junction, clone pSHI323
gi215614|gb|M13317|PP1CN23M [215614]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1035 nucleotide neighbors)

40

M13316

Bacteriophage P1 Cln recombinase activated cross over site, left junction, clone pSHI323
gi215613|gb|M13316|PP1CN23L [215613]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

45

M13315

Bacteriophage P1 Cln recombinase activated cross over site, right junction, clone pSHI322
gi215612|gb|M13315|PP1CN22R [215612]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

50

55

5

M13314

215

Bacteriophage P1 Cln recombinase activated cross over site, left junction, clone pSHI322
gi215611|gb|M13314|PP1CIN22L [215611]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1401 nucleotide neighbors)

10

M13313

Bacteriophage P1 Cln recombinase activated cross over site, right junction, clone pSHI321
gi215610|gb|M13313|PP1CIN21R [215610]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 7 nucleotide neighbors)

M13312

Bacteriophage P1 Cln recombinase activated cross over site, left junction, clone pSHI321
gi215609|gb|M13312|PP1CIN21L [215609]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1058 nucleotide neighbors)

15

M16568

Bacteriophage P1 c4 repressor gene, complete cds
gi215603|gb|M16568|PP1C4 [215603]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

20

M13326

Bacteriophage P1 Cln recombinase activated cross over site, junction III, clone pSHI326
gi215602|gb|M13326|PP1C26III [215602]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1192 nucleotide neighbors)

M13322

Bacteriophage P1 Cln recombinase activated cross over site, junction III, clone pSHI325
gi215601|gb|M13322|PP1C25III [215601]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1231 nucleotide neighbors)

25

J05651

Bacteriophage P1 modulator protein (bof) gene, complete cds
gi215598|gb|J05651|PP1BOFY1 [215598]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

30

M33224

Bacteriophage P1 regulatory protein (bof) gene, complete cds
gi215596|gb|M33224|PP1BOFFO [215596]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

35

M10288

E.coli/bacteriophage P1 loxR recombination site
gi146647|gb|M10288|ECOLOXR [146647]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 3 nucleotide neighbors)

M10289

E.coli/bacteriophage P1 loxL recombination site
gi146646|gb|M10289|ECOLOXL [146646]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

40

M10290

E.coli loxB site, which can recombine with bacteriophage P1 loxP site
gi146645|gb|M10290|ECOLOXB [146645]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

45

50

55

M10287

Bacteriophage P1 loxP X loxP recombination site
gi215627|gb|M10287|PP1LOXPX [215627]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

M74046

Bacteriophage P1 pacA and pacB genes, complete cds
gi215634|gb|M74046|PP1PACAB [215634]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

M95666

Bacteriophage P1 gene 10, doc and phd genes, complete cds
gi463276|gb|M95666|PP1PHDDOC [463276]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 4 protein links, or 1 nucleotide neighbor)

M25604

Bacteriophage Q-beta mutated autonomously replicating sequence MDV1 RNA fragment
gi556359|gb|M25604|PQBARSMT [556359]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

V00643

first half of the phage Q-beta gene for coat protein
gi15088|emb|M00643|LEQBET [15088]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

M25167

Bacteriophage Q-beta RNA fragment recovered from replicase binding complex
gi556362|gb|M25167|PQBREPLCB [556362]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

M24876

Bacteriophage Q-beta replicase RNA, 5' end
gi556360|gb|M24876|PQBREPLCA [556360]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

M25444

Synthetic bacteriophage Q-beta DNA
gi209118|gb|M25444|SYNQBTERM [209118]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

M25463

Bacteriophage Q-beta self-replicating microvariant (+) RNA
gi532489|gb|M25463|PQBMVSRNA [532489]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

M25014

Bacteriophage Q-beta RNA replicase gene, 5' end, and maturation protein gene, 3' end
gi294316|gb|M25014|PQBREPLC [294316]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

M25065

Bacteriophage Q-beta RNA sequence with putative stem loop
gi294315|gb|M25065|PQBLOOP [294315]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 nucleotide neighbors)

5

217

M10265

Bacteriophage Q-beta RNA molecule with the ability to replicate extracellularly

gi|215726|gb|M10265|PQBRNA [215726]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

10

M24815

Bacteriophage Q-beta specified replicase subunit RNA,

gi|215725|gb|M24815|PQBREPL [215725]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 nucleotide neighbors)

M25461

Bacteriophage Q-beta plus-strand RNA, 5' terminus

gi|215724|gb|M25461|PQBPSSE [215724]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

15

M25462

Bacteriophage Q-beta plus-strand RNA, 3' terminus

gi|215723|gb|M25462|PQBPSSE [215723]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 8 nucleotide neighbors)

20

M24871

Bacteriophage Q-beta nanovariant WSIII RNA

gi|215722|gb|M24871|PQBNVWSIC [215722]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

25

M24870

Bacteriophage Q-beta nanovariant WSII RNA

gi|215721|gb|M24870|PQBNVWSIB [215721]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

M24869

Bacteriophage Q-beta nanovariant WSI RNA

gi|215720|gb|M24869|PQBNVWSIA [215720]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

30

M10495

Coliphage Q-beta MDV-1(+) RNA

gi|215719|gb|M10495|PQBMDV1A [215719]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 10 nucleotide neighbors)

35

J02484

bacteriophage qbeta coat protein cistron first half

gi|215717|gb|J02484|PQBPCP5 [215717]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

M57754

Bacteriophage Q-beta minus strand RNA, 5' terminus

gi|215716|gb|M57754|PQBBSSE [215716]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 8 nucleotide neighbors)

40

M24297

Bacteriophage Q-beta 5'-terminal region of the minus strand

gi|215715|gb|M24297|PQBSEND [215715]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

45

50

55

5

M10695 218
Bacteriophage Q-beta, MDV-1 RNA
gi215714|gb|M10695|PQBI1R [215714]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 12 nucleotide neighbors)

10

M24827
Bacteriophage R17 A protein gene, 5' end
gi216078|gb|M24827|R17RNACIS|216078)
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 5 nucleotide neighbors)

15

M24829
Bacteriophage R17 coat protein gene, 5' end
gi216075|gb|M24829|R17CP5 [216075]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 5 nucleotide neighbors)

20

J02488
bacteriophage r17 rna synthetase initiation site
gi|216080|gb|J02488|17RNASYN [216080]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 2 protein links, or 6 nucleotide neighbors)

25

J02487
bacteriophage r17 coat protein initiation site
gi|216073|gb|J02487|R17COATP [216073]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

J02486
bacteriophage r17 a protein initiation site
gi|216071|gb|J02486|R17APROT [216071]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

30

M24826
Bacteriophage R17 coat protein RNA fragment
gil216077|gb|M24826|R17CPRAA [216077]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

35

M24296
Bacteriophage R17 3'-terminal fragment A RNA
gi216070|gb|M24296|R173TFA [216070]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 9 nucleotide neighbors)

TFN
structure refinement for a 24-nucleotide rna hairpin, nmr, minimized average
structure ribonucleic acid, hairpin, bacteriophage r17 mol_id: 1; molecule: r17c; chain: null; engineered: yes
gil1942336pdb[1TFN][1942336]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 structure link)

40

RPEA
rna (5'-d(ggggpgapcpuppgpapcpagpupcpapcpgp cpapgpupcpupapu-3') (24-mer rna
hairpin coat protein binding site for bacteriophage t17) (nmr, minimized average structure)
gil1421020[pdb][IRHT][1421020]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 structure link)

45

50

55

M14428

Bacteriophage S13 circular DNA, complete genome
gi216089|gb|M14428|S13CG [216089]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 12 protein links, 26 nucleotide neighbors, or 1 genome link)

J05393

Bacteriophage T1 DNA N-6-adenine-methyltransferase (M.T1) gene, complete cds
gi166163|gb|J05393|BT1NAMTA [166163]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

L46845

Bacteriophage T2 frd3, frd2 genes, complete cds

gi951387|gb|L46845|PT2FRD32G [951387]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 17 nucleotide neighbors)

L43611

Bacteriophage T2 fibrin (wac) gene, complete cds

gi903869|gb|L43611|PT2WAC [903869]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 4 nucleotide neighbors)

M24812

Bacteriophage T2 secondary structure RNA sequence

gi215796|gb|M24812|PT2RNA [215796]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 nucleotide neighbors)

M22342

Bacteriophage T2 DNA-(adenine-N6)methyltransferase (dam) gene, complete cds

gi215792|gb|M22342|PT2DAM [215792]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

S57515

orf 61.2 (intergenic region between 41 and 61) [bacteriophage T2, Genomic, 323 nt]

gi298524|gb|S57515|S57515 [298524]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

X05312

Bacteriophage T2 gene 38 for receptor recognizing protein

gi15197|emb|X05312|MYT2G38 [15197]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

X04442

Bacteriophage T2 gene 37 for receptor recognizing protein

gi15195|emb|X04442|MYT2G37 [15195]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

X12460

Bacteriophage T2 gene 32 mRNA for single-stranded DNA binding protein

gi15192|emb|X12460|MYT2G32 [15192]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 14 nucleotide neighbors)

X57797

Bacteriophage T2 gene for gp12

gi14875|emb|X56555|BT2GP12 [14875]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 2 nucleotide neighbors)

X01755

Bacteriophage T2 tail fiber gene 36

gi|15189|emb|X01755|MYT2F36 [15189]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

M14784

Bacteriophage T3 strain amNG220B right end, tail fiber protein, lysis protein and DNA packaging proteins, complete cds

gi|215810|gb|M14784|PT3RE [215810]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 9 protein links, or 10 nucleotide neighbors)

SEQ_PT3RNAPOL

Bacteriophage T3 RNA polymerase III gene, 5' end

gi|710559|gb|SEQ_PT3RNAPOL [710559]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

M22610

Bacteriophage T3 RNA polymerase III gene, 3' end

gi|340722|gb|M22610|PT3RNAPOL2 [340722]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M22609

Bacteriophage T3 RNA polymerase III gene, 5' end

gi|340721|gb|M22609|PT3RNAPOL1 [340721]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

X05031

Bacteriophage T3 gene region 1-2.5 with primary origin of replication

gi|15719|emb|X05031|POT3ORI [15719]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 11 protein links, or 5 nucleotide neighbors)

X03964

Bacteriophage T3 early control region pos. 308-810 from genome left end

gi|15718|emb|X03964|POT3EP [15718]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 20 nucleotide neighbors)

X17255

Bacteriophage T3 gene 1 to gene 11

gi|15682|emb|X17255|POT3111G [15682]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,4 MEDLINE links, 36 protein links, 17 nucleotide neighbors, or 1 genome link)

X15840

Phage T3 gene 10

gi|15625|emb|X15840|PODT3G10 [15625]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 3 nucleotide neighbors)

X02981

Bacteriophage T3 gene 1 for RNA polymerase

gi|15561|emb|X02981|PODOT3P [15561]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

J02503

bacteriophage t3 5' end, terminally redundant sequence (trs)

gi|215816|gb|J02503|PT3TRS1 [215816]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

5

221

SEG_PT3TRS

bacteriophage T3 5' end, terminally redundant sequence (trs)

gi215818|gb|SEG_PT3TRS [215818]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

10

J02504

bacteriophage T3 3' end, terminally redundant sequence (trs)

gi215817|gb|J02504|PT3TRS2 [215817]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

15

HYPERLINK <http://www.rs.noda.sut.ac.jp/~kunisawa> <http://www.rs.noda.sut.ac.jp/~kunisawa>
Bacteriophage T4 genomic database compiled by Arisaka et al.

X95646

Bacteriophage T5 DNA for region 60.5%-71% of the T5 genome

gi2791557|emb|AJ001191|BTJ001191 [2791557]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 7 MEDLINE links, 12 protein links, or 6 nucleotide neighbors)

20

X56847

Bacteriophage T5 genomic region encoding early genes D10-D15

gi15407|emb|X12930|MYT5D10 [15407]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 5 protein links, or 4 nucleotide neighbors)

25

AF039886

Bacteriophage T5 subclone T5.5.3r5.18r, single pass sequence, genomic survey sequence

gi2811154|gb|AF039886|AF039886 [2811154]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039885

Bacteriophage T5 subclone T5.40f.41f, single pass sequence, genomic survey sequence

gi2811153|gb|AF039885|AF039885 [2811153]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

30

AF039884

Bacteriophage T5 subclone T5.26.fr, single pass sequence, genomic survey sequence

gi2811152|gb|AF039884|AF039884 [2811152]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

35

AF039883

Bacteriophage T5 subclone 10-T5.5.7F, single pass sequence, genomic survey sequence

gi2811151|gb|AF039883|AF039883 [2811151]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

40

AF039882

Bacteriophage T5 subclone 41-T5.5.4BF, single pass sequence, genomic survey sequence

gi2811150|gb|AF039882|AF039882 [2811150]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

45

AF039881

Bacteriophage T5 subclone 39-T5.5.4aF, single pass sequence, genomic survey sequence

gi2811149|gb|AF039881|AF039881 [2811149]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

50

55

AF039880

Bacteriophage T5 subclone 19-T5.7.2r, single pass sequence, genomic survey sequence
gi|2811148|gb|AF039880|AF039880 [2811148]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039879

Bacteriophage T5 subclone 18-T5.7.2F, single pass sequence, genomic survey sequence
gi|2811147|gb|AF039879|AF039879 [2811147]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039878

Bacteriophage T5 subclone 11-T5.5.7R, single pass sequence, genomic survey sequence
gi|2811146|gb|AF039878|AF039878 [2811146]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 2
nucleotide neighbors)

AF039877

Bacteriophage T5 subclone T5.4FR, single pass sequence, genomic survey sequence
gi|2811145|gb|AF039877|AF039877 [2811145]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039876

Bacteriophage T5 subclone 22-T5.16R, single pass sequence, genomic survey sequence
gi|2811144|gb|AF039876|AF039876 [2811144]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039875

Bacteriophage T5 subclone 21-T5.16R, single pass sequence, genomic survey sequence
gi|2811143|gb|AF039875|AF039875 [2811143]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039874

Bacteriophage T5 subclone 21-T5.16F, single pass sequence, genomic survey sequence
gi|2811142|gb|AF039874|AF039874 [2811142]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039873

Bacteriophage T5 subclone 09-T5.6F, single pass sequence, genomic survey sequence
gi|2811141|gb|AF039873|AF039873 [2811141]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039872

Bacteriophage T5 subclone 09-T5.6R, single pass sequence, genomic survey sequence
gi|2811140|gb|AF039872|AF039872 [2811140]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 2 nucleotide neighbors)

AF039871

Bacteriophage T5 subclone 04-T5.26R, single pass sequence, genomic survey sequence
gi|2811139|gb|AF039871|AF039871 [2811139]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039870

Bacteriophage T5 subclone 13-T5.42F, single pass sequence, genomic survey sequence
gi|2811138|gb|AF039870|AF039870 [2811138]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

5

223

X69460

Bacteriophage T5 ltf gene for L-shaped tail fibers

gi|15415|emb|X69460|MYT5LTF [15415]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 1 protein link, or 4 nucleotide neighbors)

10

X03402

Bacteriophage T5 D15 gene for 5' exonuclease

gi|15413|emb|X03402|MYT5EXOG [15413]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

Z11972

Bacteriophage T5 tRNA-Tyr, tRNA-Glu, tRNA-Trp, tRNA-Phe, tRNA-Cys and

tRNA-Asn genes, and ORFs 91aa, 90aa, 42aa and 172aa

gi|15795|emb|Z11972|T56TRNAG [15795]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 3 nucleotide neighbors)

15

X03898

Bacteriophage T5 genes for tRNA-His, -Ser and -Leu

gi|15786|emb|X03898|STT5RN1 [15786]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2 MEDLINE links)

20

X04177

Bacteriophage T5 gene for transfer RNA-Gln

gi|15421|emb|X04177|MYT5TRNQ [15421]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

25

X03899

Bacteriophage T5 genes for tRNA-Val, -Lys, -Met, -Pro and -Ile3

gi|15787|emb|X03899|STT5RN2 [15787]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

X03798

Bacteriophage T5 gene for tRNA-Asp (GUC)

gi|15472|emb|X03798|NCT5TRDG [15472]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

30

Y00364

Bacteriophage T5 tRNA gene cluster (27.8%-22.4%)

gi|15420|emb|Y00364|MYT5TRN [15420]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 13 nucleotide neighbors)

35

X03140

Bacteriophage T5 DNA with rho-dependent transcription terminator (Hind III-P fragment)

gi|15417|emb|X03140|MYT5RHO [15417]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

Z35070

Bacteriophage T6 DNA

gi|535228|emb|Z35074|MYEREGBT6 [535228]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

40

45

50

55

AF060870

Coliphage T6 small subunit distal tail fiber (gene 36) gene, partial cds; and large subunit distal tail fiber (gene 37) and tail fiber adherin (gene 38) genes, complete cds
gi|3676458|gb|AF052605|AF052605 [3676458]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 protein links, or 2 nucleotide neighbors)

Z35072

Bacteriophage T6 DNA encoding ORF19.1 gene and g19 gene
gi|535232|emb|Z35072|MYTAILT6 [535232]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

X12488

Bacteriophage T6 gene 32 mRNA for single-stranded DNA binding protein
gi|15843|emb|X12488|MYT6G32 [15843]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 14 nucleotide neighbors)

Z78095

Bacteriophage T6 DNA (1506 bp)
gi|1488562|emb|Z78095|BPHZ78095 [1488562]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 4 nucleotide neighbors)

Z35079

Bacteriophage T6 DNA for lp5, lp6
gi|535215|emb|Z35079|MY57BT6 [535215]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

X68725

E.coli bacteriophage T6 gene for beta-glucosyl-HMC-alpha-glucosyl-transferase
gi|296439|emb|X68725|ECT6 [296439]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

X69894

Bacteriophage T6 alt gene for ADP-Ribosyltransferase
gi|15422|emb|X69894|MYT6ADP [15422]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

L46846

Bacteriophage T6 frd3, frd2 genes, complete cds
gi|951390|gb|L46846|PT6FRD32G [951390]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 2 protein links)

M27738

Bacteriophage T6 translational repressor protein (regA), complete cds
gi|215993|gb|M27738|PT6REGA [215993]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 5 nucleotide neighbors)

M38465

Bacteriophage T6 DNA ligase gene, complete cds
gi|215991|gb|M38465|PT6LIG55 [215991]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

V01146

Genome of bacteriophage T7

gi|431187|emb|V01146|T7CG (431187)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,13 MEDLINE links, 60 protein links, 105 nucleotide neighbors, or 1 genome link)

X60322

Bacteriophage alpha3 genes A, B, K, C, D, E, J, F, G, H

gi|14775|emb|X60322|BACALPHA (14775)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 10 protein links, 22 nucleotide neighbors, or 1 genome link)

X13332

Bacteriophage alpha3 DNA for origin of replication

gi|15093|emb|X13332|MLA3ORPL (15093)

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

X12611

Bacteriophage alpha3 gene for protein A part, finger domain

gi|15092|emb|X12611|MLA3AFIN (15092)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 6 nucleotide neighbors)

X15721

Bacteriophage alpha3 deletion mutation DNA for the origin region (-ori) of replication

gi|14774|emb|X15721|BA3DMOR9 (14774)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 11 nucleotide neighbors)

X15720

Bacteriophage alpha3 deletion mutant DNA for the origin region (-ori) of replication

gi|14773|emb|X15720|BA3DMOR8 (14773)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 nucleotide neighbor)

X15719

Bacteriophage alpha3 insertion mutant DNA for the origin region (-ori) of replication

gi|14772|emb|X15719|BA3DMOR7 (14772)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 10 nucleotide neighbors)

X15718

Bacteriophage alpha3 deletion mutation DNA for origin region (-ori) of replication

gi|14771|emb|X15718|BA3DMOR6 (14771)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 11 nucleotide neighbors)

X15717

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication

gi|14770|emb|X15717|BA3DMOR5 (14770)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 9 nucleotide neighbors)

X15716

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication

gi|14769|emb|X15716|BA3DMOR4 (14769)

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 10 nucleotide neighbors)

X15715

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication
gi|14768|emb|X15715|BA3DMOR3 [14768]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X15714

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication
gi|14767|emb|X15714|BA3DMOR2 [14767]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X15713

Bacteriophage alpha3 deletion mutant DNA for the origin region (-ori) of replication
gi|14766|emb|X15713|BA3DMOR1 [14766]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X62059

Bacteriophage alpha3 origin of cDNA synthesis (oriGA)
gi|14763|emb|X62059|AL3ORIGA [14763]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

X62058

Bacteriophage alpha3 origin of cDNA synthesis (oriAA)
gi|14762|emb|X62058|AL3ORIAA [14762]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

J02444

Bacteriophage alpha3 origin of DNA replication
gi|166103|gb|J02444|AL3ORJ [166103]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 12 nucleotide neighbors)

M25640

Bacteriophage alpha-3 H protein gene, complete cds
gi|166101|gb|M25640|AL3HP [166101]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 13 nucleotide neighbors)

M10631

Bacteriophage alpha-3 cleavage site for phage phi-X174 gene A protein
gi|166099|gb|M10631|AL3CSA [166099]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

X00774

Bacteriophage alpha-3 gene J sequence
gi|15431|emb|X00774|NCBA3J [15431]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 2 nucleotide neighbors)

M25640

Bacteriophage alpha-3 H protein gene, complete cds
gi|166101|gb|M25640|AL3HP [166101]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 13 nucleotide neighbors)

M10631

Bacteriophage alpha-3 cleavage site for phage phi-X174 gene A protein
gi|166099|gb|M10631|AL3CSA [166099]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

5

227

J02459

Bacteriophage lambda, complete genome

gi215104|gb|J02459|LAMCG [215104]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,87 MEDLINE links, 67 protein links, 190 nucleotide neighbors, or 1 genome link)

10

J02482

Bacteriophage phi-X174, complete genome

gi216019|gb|J02482|PX1CG [216019]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,23 MEDLINE links, 11 protein links, 26 nucleotide neighbors, or 1 genome link)

15

J02454

Bacteriophage G4, complete genome

gi215415|gb|J02454|PG4CG [215415]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 11 protein links, 20 nucleotide neighbors, or 1 genome link)

X60323

Bacteriophage phiK complete genome

gi1478118|emb|X60323|BPHKCG [1478118]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,10 protein links, 18 nucleotide neighbors, or 1 genome link)

20

L42820

Bacteriophage BF23 tail protein (hrs) gene, complete cds

gi1048680|gb|L42820|BBFHRS [1048680]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

25

X54455

Bacteriophage BF23 gene 17 and gene 18

gi14797|emb|X54455|BF231718G [14797]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 2 nucleotide neighbors)

30

M37097

Bacteriophage BF23 DNA, right end of terminal repetition

gi166115|gb|M37097|BBFRIGH [166115]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

M37096

Bacteriophage BF23 DNA, left end of terminal repetition

gi166114|gb|M37096|BBFLEFT [166114]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 nucleotide neighbor)

35

M37095

Bacteriophage BF23 A2-A3 gene, complete cds, and A1 gene, 5' end

gi166110|gb|M37095|BBFA2A3 [166110]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 3 protein links, or 1 nucleotide neighbor)

40

AF056281

Bacteriophage BF23 clone bf23.mac5/6.1, genomic survey sequence

gi3090930|gb|AF056281|AF056281 [3090930]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

45

50

55

AF056280

Bacteriophage BF23 clone bf23.mac3, genomic survey sequence
gi|3090929|gb|AF056280|AF056280 [3090929]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056279

Bacteriophage BF23 clone bf23.mac18/21.34, genomic survey sequence
gi|3090928|gb|AF056279|AF056279 [3090928]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056278

Bacteriophage BF23 clone bf23.mac16/19.33, genomic survey sequence
gi|3090927|gb|AF056278|AF056278 [3090927]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056277

Bacteriophage BF23 clone bf23.mac16/19.33, genomic survey sequence
gi|3090926|gb|AF056277|AF056277 [3090926]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056276

Bacteriophage BF23 clone bf23.mac12/9-9, genomic survey sequence
gi|3090925|gb|AF056276|AF056276 [3090925]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056275

Bacteriophage BF23 clone bf23.mac11/14-24, genomic survey sequence
gi|3090924|gb|AF056275|AF056275 [3090924]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056274

Bacteriophage BF23 clone bf23.57r64r, genomic survey sequence
gi|3090923|gb|AF056274|AF056274 [3090923]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 3 nucleotide neighbors)

AF056273

Bacteriophage BF23 clone bf23.54E, genomic survey sequence
gi|3090922|gb|AF056273|AF056273 [3090922]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056272

Bacteriophage BF23 clone bf23.47f.mac10/7, genomic survey sequence
gi|3090921|gb|AF056272|AF056272 [3090921]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056271

Bacteriophage BF23 clone bf23.23.66r, genomic survey sequence
gi|3090920|gb|AF056271|AF056271 [3090920]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056270

Bacteriophage BF23 clone bf23.23.64f, genomic survey sequence
gi|3090919|gb|AF056270|AF056270 [3090919]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

5

229

AF056269

Bacteriophage BF23 clone bF23.23.60r, genomic survey sequence
gi|3090918|gb|AF056269|AF056269 [3090918]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

10

AF056268

Bacteriophage BF23 clone bF23.23.60f, genomic survey sequence
gi|3090917|gb|AF056268|AF056268 [3090917]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

15

AF056267

Bacteriophage BF23 clone bF23.23.59r, genomic survey sequence
gi|3090916|gb|AF056267|AF056267 [3090916]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

20

AF056266

Bacteriophage BF23 clone bF23.23.59f, genomic survey sequence
gi|3090915|gb|AF056266|AF056266 [3090915]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

25

AF056265

Bacteriophage BF23 clone bF23.23.56r, genomic survey sequence
gi|3090914|gb|AF056265|AF056265 [3090914]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

30

AF056264

Bacteriophage BF23 clone bF23.23.56f, genomic survey sequence
gi|3090913|gb|AF056264|AF056264 [3090913]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056263

Bacteriophage BF23 clone bF23.23.68f5r, genomic survey sequence
gi|3090912|gb|AF056263|AF056263 [3090912]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

35

AF056262

Bacteriophage BF23 clone bF23.23.43f.66f, genomic survey sequence
gi|3090911|gb|AF056262|AF056262 [3090911]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

40

AF056261

Bacteriophage BF23 clone bF23.23.2fr, genomic survey sequence
gi|3090910|gb|AF056261|AF056261 [3090910]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056260

Bacteriophage BF23 clone bF23.23.55f, genomic survey sequence
gi|3090909|gb|AF056260|AF056260 [3090909]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

45

AF056259

Bacteriophage BF23 clone bF23.23.53r, genomic survey sequence
gi|3090908|gb|AF056259|AF056259 [3090908]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

50

55

5

230

AF056258

Bacteriophage BF23 clone bf23.23.53.f, genomic survey sequence
gi|3090907|gb|AF056258|AF056258 [3090907]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

10

AF056257

Bacteriophage BF23 clone bf23.23.52.r, genomic survey sequence
gi|3090906|gb|AF056257|AF056257 [3090906]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

15

AF056256

Bacteriophage BF23 clone bf23.23.52.f, genomic survey sequence
gi|3090905|gb|AF056256|AF056256 [3090905]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

20

AF056255

Bacteriophage BF23 clone bf23.23.49.r, genomic survey sequence
gi|3090904|gb|AF056255|AF056255 [3090904]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

25

AF056254

Bacteriophage BF23 clone bf23.23.49.f, genomic survey sequence
gi|3090903|gb|AF056254|AF056254 [3090903]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

30

AF056253

Bacteriophage BF23 clone bf23.23.48.r, genomic survey sequence
gi|3090902|gb|AF056253|AF056253 [3090902]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056252

Bacteriophage BF23 clone bf23.23.48.f, genomic survey sequence
gi|3090901|gb|AF056252|AF056252 [3090901]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

35

AF056251

Bacteriophage BF23 clone bf23.23.44.r, genomic survey sequence
gi|3090900|gb|AF056251|AF056251 [3090900]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

40

AF056250

Bacteriophage BF23 clone bf23.23.41.f, genomic survey sequence
gi|3090899|gb|AF056250|AF056250 [3090899]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056249

Bacteriophage BF23 clone bf23.23.22.a.r, genomic survey sequence
gi|3090898|gb|AF056249|AF056249 [3090898]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

45

AF056248

Bacteriophage BF23 clone bf23.23.22.a.f, genomic survey sequence
gi|3090897|gb|AF056248|AF056248 [3090897]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

50

55

5

231

AF056247

Bacteriophage BF23 clone bf23.23.68.r, genomic survey sequence
gi|3090896|gb|AF056247|AF056247 [3090896]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

10

Z50114

Bacteriophage BF23 DNA for putative tail protein gene
gi|2464952|emb|Z50114|BF23LATE [2464952]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 protein link)

D12824

15

Bacteriophage BF23 genes for minor tail protein gp24 and major tail protein gp25, complete cds
gi|520578|dbj|D12824|BBF2TAIL [520578]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

Z34953

Bacteriophage K3 ip9, ip7 and ip8 genes
gi|535261|emb|Z34953|MYK3IP978 [535261]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

20

Z35075

Bacteriophage K3 DNA for Ip3 and Ip4
gi|535229|emb|Z35075|MYEORF64K [535229]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

X05560

25

Bacteriophage K3 gene 38 for receptor recognizing protein
gi|151112|emb|X05560|MYK3G38 [151112]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

X04747

30

Bacteriophage K3 gene 37 for receptor recognizing protein
gi|151110|emb|X04747|MYK3G37 [151110]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

X01754

Bacteriophage K3 tail fiber gene 36
gi|15108|emb|X01754|MYK3F36 [15108]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

35

M16812

Bacteriophage K3 'r' lysis gene, complete cds
gi|215503|gb|M16812|PK3LYST [215503]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

L46833

40

Bacteriophage K3 frd3, frd2 genes, complete cds
gi|951377|gb|L46833|PK3FRD32G [951377]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 2 nucleotide neighbors)

L43613

45

Bacteriophage K3 fibrin (wac) gene, complete cds
gi|903861|gb|L43613|PK3WAC [903861]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 4 nucleotide neighbors)

50

55

5

X01753

Bacteriophage OX2 tail fiber gene 36

gi115122|emb|X01753|MYOX2F36 [15122]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

10

L43612

Bacteriophage OX2 fibrin (wac) gene, complete cds

gi1903848|gb|L43612|OX2WAC [903848]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 4 nucleotide neighbors)

Z46880

Bacteriophage OX2 stp gene

gi1599663|emb|Z46880|BPOX2STP [599663]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

15

X05675

Bacteriophage OX2 gene 38 for receptor-recognizing protein and flanking regions

gi115124|emb|X05675|MYOX2G38 [15124]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

20

M33533

Bacteriophage RB18 translational repressor protein (regA) and Orf43.1, complete cds

gi216083|gb|M33533|RB18REGA [216083]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

25

AF033329

Bacteriophage RB18 single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi12645788|gb|AF033329|AF033329 [2645788]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 11 nucleotide neighbors)

M86231

Bacteriophage RB69 gene 62, 3' end; RegA (regA) gene, complete cds

gi1215354|gb|M86231|P6962REGA [215354]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

30

AF033332

Bacteriophage RB69 single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi12645794|gb|AF033332|AF033332 [2645794]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 12 nucleotide neighbors)

35

U34036

Bacteriophage RB69 DNA polymerase (43) gene, complete cds

gi1237125|gb|U34036|BRU34036 [1237125]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

V01145

Bacteriophage H1 genome fragment Each Thymine given in this sequence represents a HMU-residue
(HMU = 5-hydroxymethyluracil)

gi115557|emb|V01145|PODOH1 [15557]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

40

X05676

Bacteriophage M1 gene 38 for receptor recognizing protein and flanking regions

gi115114|emb|X05676|MYM1G38 [15114]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

45

50

55

5

10

15

20

25

30

35

40

45

50

55

AF034575

Bacteriophage M1 putative integrase (int) gene, complete cds, and attP region, complete sequence
gi|2662472|gb|AF034575|AF034575 [2662472]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

AF033321

Bacteriophage M1 single-stranded binding protein (gene 32) gene, partial cds, and 5' region
gi|2645772|gb|AF033321|AF033321 [2645772]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 17 nucleotide neighbors)

X55190

Bacteriophage Tu1a 37 and 38 genes for receptor-recognizing proteins 37 and 38 (respectively), partial cds
gi|14860|emb|X55190|BPTU1A [14860]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

AF033334

Bacteriophage Tu1b single-stranded binding protein (gene 32) gene, partial cds, and 5' region
gi|2645798|gb|AF033334|AF033334 [2645798]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 5 nucleotide neighbors)

X55191

Bacteriophage Tu1b 37 gene for receptor-recognizing protein 37 (partial cds), 38 gene for receptor-recognizing protein 38, and t gene (partial cds)

gi|14863|emb|X55191|BPTU1B [14863]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 3 nucleotide neighbors)

X13065

Bacteriophage phi80 early region

gi|14800|emb|X13065|BP80ER [14800]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 8 protein links, or 6 nucleotide neighbors)

D00360

Bacteriophage phi80 cor gene

gi|217782|dbj|D00360|P8080COR [217782]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 protein link)

X01639

Bacteriophage phi 80 DNA-fragment with replication origin

gi|15828|emb|X01639|XXPHI80 [15828]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 25 nucleotide neighbors)

X04051

Lambdoid bacteriophage phi 80 int-xis region (integrase-excisionase region)

gi|15770|emb|X04051|STPHI80X [15770]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

X06751

Phage Phi80 DNA for major coat protein

gi|15768|emb|X06751|STPHI80C [15768]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 11 nucleotide neighbors)

X75949

Bacteriophage phi80 DNA for ORF x171.5 and ORF x171.28

gi|458811|emb|X75949|ECORF171B [458811]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 28 nucleotide neighbors)

L40418

Bacteriophage phi-80 gene, complete cds

gi|1019107|gb|L40418|P80A [1019107]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

M24831

Bacteriophage phi-80 Tyr-tRNA gene, 3' end

gi|215363|gb|M24831|P80TGY [215363]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 43 nucleotide neighbors)

M10670

Bacteriophage phi-80 replication origin

gi|215361|gb|M10670|P80ORI [215361]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

M24825

Bacteriophage phi-80 RNA fragment

gi|215360|gb|M24825|P80MJA [215360]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M11919

Bacteriophage phi-80 cI immunity region encoding the N gene

gi|215358|gb|M11919|P80CI [215358]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

M10891

Bacteriophage phi-80 attP site DNA

gi|215357|gb|M10891|P80ATTP [215357]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M19473

Bacteriophage 933J (from E.coli) proviral Shiga-like toxin type I subunits A and B genes, complete cds

gi|215072|gb|M19473|J93SLTI [215072]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 2 protein links, or 20 nucleotide neighbors)

Y10775

Bacteriophage 933W ileX, stx2A and stx2B genes

gi|193820|emb|Y10775|BP933ILEX [1938206]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 36 nucleotide neighbors)

X83722

Bacteriophage 933W slt-II gene

gi|1490229|emb|X83722|B933WSLT [1490229]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 20 nucleotide neighbors)

X07865

Bacteriophage 933W slt-II gene for Shiga-like toxin type II subunit A and B

gi|14892|emb|X07865|BWSLTII [14892]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 29 nucleotide neighbors)

M16625

Bacteriophage H19B (from E.coli) sltA and sltB genes encoding Shiga-like toxin I subunits A and B, complete cds

gi|215043|gb|M16625|H19BSLT [215043]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 24 nucleotide neighbors)

5

M17358

235

Bacteriophage H19B shiga-like toxin-1 (SLT-1) A and B subunit DNA, complete cds

gi215046|gb|M17358|H19BSLTA [215046]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 20 nucleotide neighbors)

10

U29728

Bacteriophage N4 single-stranded DNA-binding protein (N4SSB) gene, complete cds

gi939708|gb|U29728|BNU29728 [939708]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 1 protein link)

J02580

Bacteriophage PA-2 (E.coli porcine strain isolate) Rz gene, 5'end; ORF2, outer membrane porin protein (ic) and ORF1 genes, complete cds

gi215366|gb|J02580|PA2LC [215366]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 4 nucleotide neighbors)

15

U32222

Bacteriophage 186, complete sequence

gi3337249|gb|U32222|B1U32222 [3337249]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 46 protein links, or 5 nucleotide neighbors)

20

X51522

Bacteriophage P4 complete DNA genome

gi450916|emb|X51522|MYP4CG [450916]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 13 protein links, 6 nucleotide neighbors, or 1 genome link)

25

X92588

Bacteriophage 82 orf33, orf151, orf56, orf96, rus, orf45, and Q genes

gi1051111|emb|X92588|BAC82HOLL [1051111]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,7 protein links, or 1 nucleotide neighbor)

J02803

Bacteriophage 82 antitermination protein (Q) gene, complete cds

gi215364|gb|J02803|P82Q [215364]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

30

U02466

Bacteriophage HK022 (cro), (cII) and (O) genes, complete cds, (P) gene, partial cds

gi407285|gb|U02466|BHU02466 [407285]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 5 protein links, or 1 nucleotide neighbor)

35

M26291

Bacteriophage D108 regulatory DNA-binding protein (ner) gene, complete cds

gi166194|gb|M26291|D18NER [166194]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

40

M11272

Bacteriophage D108 left-end DNA

gi166193|gb|M11272|D18LEDNA [166193]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

M18902

Bacteriophage D108 kil gene encoding a replication protein, 3' end; and containing three ORFs, complete cds

gi166191|gb|M18902|D18KIL [166191]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

45

50

55

M10191

Bacteriophage D108, left end with Mu A protein binding sites L1 and L2

gi|166190|gb|M10191|D18BSL [166190]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 5 nucleotide neighbors)

J02447

bacteriophage d108 gene a 5' end

gi|166189|gb|J02447|D18AAA [166189]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

V00865

Bacteriophage D108 fragment from genes A and ner (C-terminus of ner and N-terminus of A)

gi|15437|emb|V00865|NCD108 [15437]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

X01914

Bacteriophage IKe gene for DNA binding protein

gi|14957|emb|X01914|INIKEDBP [14957]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

AF064539

Bacteriophage N15, complete genome

gi|192683|gb|AF064539|AF064539 [3192683]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 60 protein links, 26 nucleotide neighbors, or 1 genome link)

U02303

Bacteriophage IPhi. complete genome

gi|3676280|gb|U02303|B2U02303 [3676280]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 10 protein links, or 1 genome link)

AF007792

Bacteriophage Mu late morphogenetic region

gi|3551775|gb|AF007792|AF007792 [3551775]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

U24159

Bacteriophage HP1 strain HP1c1, complete genome

gi|1046235|gb|U24159|BHU24159 [1046235]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 6 MEDLINE links, 41 protein links, 8 nucleotide neighbors, or 1 genome link)

Z71579

Bacteriophage S2 type A 5.6 kb DNA fragment

gi|1679806|emb|Z71579|BPHS1ADNA [1679806]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 MEDLINE links, 9 protein links, or 9 nucleotide neighbors)

X53238

Klebsiella sp. bacteriophage K11 gene 1 for RNA polymerase

gi|14984|emb|X53238|KSK11RPO [14984]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

5

X85010

Bacteriophage A511 plyS11 gene

gi|853748|emb|X85010|BPA511PLY [853748]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

10

U29728

Bacteriophage N4 single-stranded DNA-binding protein (N4SSB) gene, complete cds

gi|939708|gb|U29728|BNU29728 [939708]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, or 1 protein link)

15

J02445

bacteriophage bo1 3'-terminal region rna

gi|166152|gb|J02445|BO1TR3 [166152]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 5 nucleotide neighbors)

L06183

Bacteriophage L5 (from *Leuconostoc oenos*) genome

gi|289353|gb|L06183|BL5GENM [289353]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 genome link)

20

AF074945

Mycoplasma arthritis bacteriophage MAV1, complete genome

gi|311243|gb|AF074945|AF074945 [311243]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 15 protein links, 3 nucleotide neighbors, or 1 genome link)

25

L13696

Bacteriophage L2 (from *Mycoplasma*), complete genome

gi|289338|gb|L13696|BL2CG [289338]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 MEDLINE links, 14 protein links, or 1 genome link)

X80191

Bacteriophage PP7 mRNA for maturation, coat, lysis and replicase proteins

gi|517237|emb|X80191|BPP7PR [517237]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 4 protein links, or 1 genome link)

30

M19377

Bacteriophage PF3 from *Pseudomonas aeruginosa* (New York strain), complete genome

gi|215380|gb|M19377|PF3COMNY [215380]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 9 protein links, or 5 nucleotide neighbors)

35

M11912

Bacteriophage PF3 from *Pseudomonas aeruginosa* (Nijmegen strain), complete genome

gi|215371|gb|M11912|PF3COMN [215371]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 9 protein links, 5 nucleotide neighbors, or 1 genome link)

40

V00605

Bacteriophage Pf1 gene encoding DNA binding protein

gi|14970|emb|V00605|INOPF1 [14970]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 1 nucleotide neighbor)

L05626

Bacteriophage PR4 capsid protein (P6) gene, complete cds

gi|215735|gb|L05626|PR4P6MAJA [215735]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

45

50

55

D13409

Bacteriophage phiCTX (isolated from *Pseudomonas aeruginosa*) cosR, attP, int genes

gi|217776|dbj|D13409|BPHCOSR [217776]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 3 nucleotide neighbors)

D13408

Bacteriophage phiCTX (isolated from *Pseudomonas aeruginosa*) cosL, ctx genes

gi|217775|dbj|D13408|BPHCOSLCTX [217775]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, or 3 nucleotide neighbors)

M24832

Bacteriophage f2 coat protein gene, partial cds

gi|166228|gb|M24832|F2CRNACA [166228]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

S72011

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618967|gb|AF017629|AF017629 [2618967]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017628

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618964|gb|AF017628|AF017628 [2618964]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017627

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618961|gb|AF017627|AF017627 [2618961]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017626

Bacteriophage 21 isocitrate dehydrogenase (icd) gene, partial cds; and integrase (int) gene, partial cds

gi|2618958|gb|AF017626|AF017626 [2618958]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 49 nucleotide neighbors)

AF017625

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618955|gb|AF017625|AF017625 [2618955]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017624

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618952|gb|AF017624|AF017624 [2618952]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017623

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618949|gb|AF017623|AF017623 [2618949]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017622

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618946|gb|AF017622|AF017622 [2618946]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

5

239

AF017621

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi2618943|gb|AF017621|AF017621 [2618943]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

10

D26449

Bacteriophage PS17 FI gene for tail sheath protein (gpFI) and FII gene for tail tube protein (gpFII), complete cds
gi452162|dbj|D26449|BPSFIFII [452162]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2 protein links)

15

X87627

Bacteriophage D3112 A and B genes
gi974768|emb|X87627|BPD3112AB [974768]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

20

U32623

Bacteriophage D3 transcriptional activator CII (cII) gene, complete cds
gi984852|gb|U32623|BDU32623 [984852]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 1 nucleotide neighbor)

25

L34781

Bacteriophage phi 11 holin homologue (ORF3) gene; complete cds and peptidoglycan hydrolase (lytA) gene, partial cds
gi511838|gb|L34781|BPHOLIN [511838]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 2 nucleotide neighbors)

L14810

Bacteriophage P22 (gp10) gene, complete cds, and (gp26) gene, complete cds
gi294053|gb|L14810|P22GP1026X [294053]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

30

X87420

Bacteriophage ES18 genes 24, c2, cro, c1, 18, and oL and oR operators
gi1143407|emb|X87420|BPES18GEN [1143407]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,5 protein links, or 9 nucleotide neighbors)

L42820

Bacteriophage BF23 tail protein (hrs) gene, complete cds
gi1048680|gb|L42820|BBFHRS [1048680]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

35

X14980

Bacteriophage PRD1 XV gene for protein P15 (lytic enzyme)
gi15802|emb|X14980|TEPRD1XV [15802]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

40

X06321

Bacteriophage PRD1 gene 8 for DNA terminal protein
gi15800|emb|X06321|TEPRD18 [15800]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 10 nucleotide neighbors)

45

X14336

Filamentous Bacteriophage I2-2 genome
gi14920|emb|X14336|INBI22 [14920]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 9 protein links, 1 nucleotide neighbor, or 1 genome link)

50

55

5

L05001 240
Bacteriophage X glucosyl transferase gene, complete cds
gi|216044|gb|L05001|PXFCLUSYLT [216044]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

10

M29479
Bacteriophage p4 sid and psu genes partial cds, and delta gene, complete cds gi|215701|
gb|M29479|PP4SDP [215701]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 protein links, or 4 nucleotide neighbors)

15

SEG_PP4PSUSID
Bacteriophage P4 capsid size determination protein (sid) gene, 5' end
gi|215698|gb|SEG_PP4PSUSID [215698]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

20

M29650
Bacteriophage P4 polarity suppression protein (psu) gene, complete cds
gi|215697|gb|M29650|PP4PSUSID2 [215697]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

25

M29651
Bacteriophage P4 capsid size determination protein (sid) gene, 5' end
gi|215696|gb|M29651|PP4PSUSID1 [215696]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M27748
Bacteriophage P4 gop, beta, and cII genes, complete cds and int gene, 3' end
gi|215691|gb|M27748|PP4GOPBC [215691]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 1 nucleotide neighbor)

30

K02750
Bacteriophage IKe, complete genome
gi|215061|gb|K02750|IKECG [215061]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 10 protein links, 4 nucleotide neighbors, or 1 genome link)

35

L40418
Bacteriophage phi-80 gene, complete cds
gi|1019107|gb|L40418|P80A [1019107]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

40

AF032122
Bacteriophage SII integrase (int) gene, partial cds; and bactoprenol glucosyl transferase (bgt), and glucosyl transferase II (gtII) genes, complete cds
gi|2463412|gb|AF021347|AF021347 [2463412]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 2 nucleotide neighbors)

45

M35825
Bacteriophage SF6 fragment D lysozyme gene, complete cds
gi|216105|gb|M35825|SF6LYZ [216105]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 protein link)

50

Z35479
Bacteriophage C16 ipl gene
gi|534936|emb|Z35479|BC16IP1 [534936]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

55

5

241

X12638

Bacteriophage 21 DNA for gene 2

gi|296141|emb|X12638|B21GENE2 [296141]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

10

X02501

Bacteriophage 21 DNA for left end sequence with genes 1 and 2

gi|15825|emb|X02501|XCPHA21 [15825]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

M65239

Bacteriophage 21 lysis genes S, R, and Rz, complete cds

gi|215466|gb|M65239|PH2LYSGEN [215466]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

15

M58702

Bacteriophage 21 late gene regulatory region

gi|215465|gb|M58702|PH2LATEGE [215465]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

20

M81255

Bacteriophage 21 head gene operon

gi|215454|gb|M81255|PH2HEADTL [215454]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 10 protein links, or 4 nucleotide neighbors)

25

M23775

Bacteriophage 21 glycoprotein 1 gene, complete cds, and glycoprotein gene, 5' end

gi|215451|gb|M23775|PH2GPA [215451]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

M61865

Bacteriophage 21 excisionase (xis), integrase (int) and isocitrate dehydrogenase (icd), complete cds

gi|215448|gb|M61865|PH22XISAA [215448]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 9 nucleotide neighbors)

30

S72011

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618967|gb|AF017629|AF017629 [2618967]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

35

AF017628

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618964|gb|AF017628|AF017628 [2618964]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017627

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618961|gb|AF017627|AF017627 [2618961]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

40

AF017626

Bacteriophage 21 isocitrate dehydrogenase (icd) gene, partial cds; and integrase (int) gene, partial cds

gi|2618958|gb|AF017626|AF017626 [2618958]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 49 nucleotide neighbors)

45

50

55

5

AF017625

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi2618955|gb|AF017625|AF017625 [2618955]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

10

AF017624

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi2618952|gb|AF017624|AF017624 [2618952]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

15

AF017623

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi2618949|gb|AF017623|AF017623 [2618949]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

20

AF017622

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi2618946|gb|AF017622|AF017622 [2618946]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017621

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi2618943|gb|AF017621|AF017621 [2618943]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

25

M57455

Bacteriophage 42D (clone pDB17) (from *Staphylococcus aureus*) staphylokinase gene, complete cds
gi215344|gb|M57455|P42STK [215344]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 9 nucleotide neighbors)

30

Y12633

Bacteriophage 85 DNA, promoter sequence of unknown gene
gi2058285|emb|Y12633|B85PROM [2058285]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

35

X98146

Bacteriophage P1 DNA sequence around the Op88 operator
gi1359513|emb|X98146|BP1OP88OP [1359513]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

Y07739

Staphylococcus phage Twort hoTW, plyTW genes
gi2764979|emb|Y07739|BPTWGHOLG [2764979]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 2 protein links)

40

L07580

Bacteriophage phi-11 rimA and rinB genes, required for the activation of *Staphylococcal* phage phi-11 int expression
gi166160|gb|L07580|BPHRJNAB [166160]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

45

M34832

Bacteriophage phi-11 integrase (int) and excisionase (xis) genes, complete cds
gi166157|gb|M34832|BPHTXIS [166157]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

50

55

5

M20394

Bacteriophage phi-11 S.aureus attachment site (attP)

gi166156|gb|M20394|BPHATT1 [166156]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 nucleotide neighbors)

10

X23128

Bacteriophage phi-13 integrase gene

gi1758228|emb|X82312|PHI13|INT [758228]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 3 nucleotide neighbors)

15

X61719

S.aureus phi-13 lysogen right chromosome/bacteriophage DNA junction

gi146624|emb|X61719|SAP13RJNC [46624]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

X61718

S.aureus phi-13 lysogen left chromosomal/bacteriophage DNA junction

gi146624|emb|X61718|SAP13LJNC [46624]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

20

X61717

Bacteriophage phi-13 core sequence for attachment

gi147999|emb|X61717|BP13ATT1 [14799]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, or 3 nucleotide neighbors)

25

U01875

Bacteriophage phi-13 putative regulatory region and integrase (int) gene, partial cds

gi1437118|gb|U01875|U01875 [437118]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 MEDLINE links, or 4 nucleotide neighbors)

X67739

S.aureus Bacteriophage phi-42 attP gene

gi148099|emb|X67739|BPATT1P [14809]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 nucleotide neighbors)

30

U01872

Bacteriophage phi-42 integrase (int) gene, complete cds

gi1437115|gb|U01872|U01872 [437115]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 MEDLINE links, 2 protein links, or 3 nucleotide neighbors)

35

X94423

Staphylococcus aureus bacteriophage phi-42 DNA with ORFs (restriction modification system)

gi1771597|emb|X94423|SARMS [1771597]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 1 nucleotide neighbor)

M27965

Bacteriophage L54a (from S.aureus) int and xis genes, complete cds

gi1215096|gb|M27965|L54INTXIS [215096]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, MEDLINE 1 link, 2 protein links, or 3 nucleotide neighbors)

40

U72397

Bacteriophage 80 alpha holin and amidase genes, complete cds

gi1763241|gb|U72397|B8U72397 [1763241]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 2 nucleotide neighbors)

45

50

55

5

244

AB009866

Bacteriophage phi PVL proviral DNA, complete sequence

gi|3341907|dbj|AB009866|AB009866 [3341907]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,63 protein links, or 1 nucleotide neighbor)

10

Z47794

Bacteriophage Cp-1 DNA, complete genome

gi|2288892|emb|Z47794|BPCP1XX [2288892]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 28 protein links, 1 nucleotide neighbor, or 1 genome link)

15

SEG_CP7RSIT

Bacteriophage Cp-7 (S.pneumoniae) 5' inverted terminal repeat

gi|166186|gb|SEG_CP7RSIT [166186]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

M11635

Bacteriophage Cp-7 (S.pneumoniae) DNA, 3' inverted terminal repeat

gi|166185|gb|M11635|CP7RSIT2 [166185]

20

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M11636

Bacteriophage Cp-7 (S.pneumoniae) 5' inverted terminal repeat

gi|166184|gb|M11636|CP7RSIT1 [166184]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

25

SEG_CP5RSIT

Bacteriophage Cp-5 (S.pneumoniae), 5' inverted terminal repeat

gi|166181|gb|SEG_CP5RSIT [166181]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

M11633

Bacteriophage Cp-5 (S.pneumoniae) 3' inverted terminal repeat

gi|166180|gb|M11633|CP5RSIT2 [166180]

30

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M11634

Bacteriophage Cp-5 (S.pneumoniae), 5' inverted terminal repeat

gi|166179|gb|M11634|CP5RSIT1 [166179]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

35

M34780

Bacteriophage Cp-9 muramidase (cp19) gene

gi|166187|gb|M34780|CP9CPL [166187]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

40

M34652

Bacteriophage HB-3 amidase (hbl) gene, complete cds

gi|215055|gb|M34652|HB3HBLA [215055]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

U64984

Streptococcus pyogenes phage T12 repressor, excisionase (xis), integrase(int) and erythrogenic toxin A precursor (speA) genes,

complete cds gi|1877426|gb|U64984|SPU40453 [1877426]

45

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 4 protein links, or 22 nucleotide neighbors)

50

55

5

245

X12375

Phage CP-T1 (*Vibrio cholerae*) DNA for packaging signal (pac site)

gi|15435|emb|X12375|NCCPPAC [15435]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

10

AF087814

Vibrio cholerae filamentous bacteriophage fs-2 DNA, complete genome sequence

gi|3702207|dbj|AB002632|AB002632 [3702207]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 9 protein links, or 1 genome link)

D83518

Bacteriophage KVP40 gene for major capsid protein precursor, complete cds

gi|3046858|dbj|D83518|D83518 [3046858]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

15

AF033322

Bacteriophage PST single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi|2645774|gb|AF033322|AF033322 [2645774]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 17 nucleotide neighbors)

20

X94331

Bacteriophage L cro, 24, c2, and c1 genes

gi|1469213|emb|X94331|BLCRO24C [1469213]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 protein links)

25

U82619

Shigella flexneri bacteriophage V glucosyl transferase (gtv), integrase (int) and excisionase (xis) genes, complete cds

gi|2465470|gb|U82619|SFU82619 [2465470]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 8 protein links, or 1 nucleotide neighbor)

30

35

40

45

50

55

246
Table 12

NCBI *Entrez* Nucleotide QUERY

Key words: bacteriophage and lysis

56 citations found (all selected)

AJ011581

Bacteriophage PS119 lysis genes 13, 19, 15, and packaging gene 3, complete cds
gi3676084|emb|AJ011581|BPS011581 [3676084]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 protein links, or 1 nucleotide neighbor)

AJ011580

Bacteriophage PS34 lysis genes 13, 19, 15, antiterminator gene 23, and packaging gene 3, complete cds
gi3676078|emb|AJ011580|BPS011580 [3676078]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 5 protein links, or 2 nucleotide neighbors)

AJ011579

Bacteriophage PS3 lysis genes 13, 19, 15, and packaging gene 3
gi3676073|emb|AJ011579|BPS011579 [3676073]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 protein links, or 1 nucleotide neighbor)

AF034975

Bacteriophage H-19B essential recombination function protein (erf), kil protein (kil), regulatory protein cIII (cIII), protein gp17 (17), N protein (N), cl protein (cl), cro protein (cro), cII protein (cII), O protein (O), P protein (P), ren protein (ren), Roi (roi), Q protein (Q), Shiga-like toxin A (slt-IA) and B (slt-IB) subunits, and putative holin protein (S) genes, complete cds
gi2668751|gb|AF034975 [2668751]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 20 protein links, or 30 nucleotide neighbors)

U37314

Bacteriophage lambda Rz1 protein precursor (Rz1) gene, complete cds
gi1017780|gb|U37314|BLU37314 [1017780]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 1 protein link, or 9 nucleotide neighbors)

U00005

E. coli hflA locus encoding the hflX, hflK and hflC genes, hflq gene, complete cds; miaA gene, partial cds
gi436153|gb|U00005|ECOHFLA [436153]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 MEDLINE

5

247

links, 5 protein links, or 8 nucleotide neighbors)

10

U32222

Bacteriophage 186, complete sequence
gi3337249|gb|U32222|B|U32222 [3337249]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE
links, 46 protein links, or 5 nucleotide neighbors)

15

AF064539

Bacteriophage N15, complete genome
gi3192683|gb|AF064539|AF064539 [3192683]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE
links, 60 protein links, 26 nucleotide neighbors, or 1 genome link)

20

AF063097

Bacteriophage P2, complete genome
gi3139086|gb|AF063097|AF063097 [3139086]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,21 MEDLINE
links, 42 protein links, 3 nucleotide neighbors, or 1 genome link)

25

Z97974

Bacteriophage phiadh lys, hol, intG, rad, and tec genes
gi2707950|emb|Z97974|BPHIADH [2707950]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE
links, 9 protein links, or 1 nucleotide neighbor)

30

AF059243

Bacteriophage NL95, complete genome
gi3088545|gb|AF059243|AF059243 [3088545]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE
links, 4 protein links, 3 nucleotide neighbors, or 1 genome link)

35

AF052431

Bacteriophage M11 A-protein, coat protein, A1-protein, and replicase
genes, complete cds
gi2981208|gb|AF052431| [2981208]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE
links, 4 protein links, or 8 nucleotide neighbors)

40

Y07739

Staphylococcus phage Twort holTW, plyTW genes
gi2764979|emb|Y07739|BFTWGHOLG [2764979]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2
protein links)

45

X94331

50

55

5

248

10

Bacteriophage L cro, 24, c2, and c1 genes
gil1469213|emblX94331|BLCRO24C [1469213]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 protein links)

X78410

15

Bacteriophage phiadh holin and lysis genes
gil793848|emblX78410|LGHOLLYS [793848]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

X99260

20

Bacteriophage B103 genomic sequence
gil1429229|emblX99260|BB103G [1429229]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 17 protein links, or 12 nucleotide neighbors)

AJ000741

25

Bacteriophage P1 darA operon
gil2462938|emblAJ000741|BPJA7641 [2462938]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 10 protein links, or 31 nucleotide neighbors)

X87420

30

Bacteriophage ES18 genes 24, c2, cro, c1, 18, and oL and oR operators
gil1143407|emblX87420|BPES18GEN [1143407]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 5 protein links, or 9 nucleotide neighbors)

L35561

35

Bacteriophage phi-105 ORFs 1-3
gil532218|gilL35561|PH5ORFHTR [532218]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 protein links)

D10027

40

Group II RNA coliphage GA genome
gil217784|dbjD10027|PGAXX [217784]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, 5 nucleotide neighbors, or 1 genome link)

V01128

45

Bacteriophage phi-X174 (cs70 mutation) complete genome
gil15535|emblV01128|PHIX174 [15535]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 MEDLINE links, 11 protein links, or 26 nucleotide neighbors)

50

55

S81763

coat gene...replicase gene (bacteriophage KU1, host=Escherichia coli, group II RNA phage, Genomic RNA, 3 genes, 120 nt)
gil1438766|gb|S81763|S81763 [1438766]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

U38906

Bacteriophage r1t integrase, repressor protein (rro), dUTPase, holin and lysin genes, complete cds
gil1353517|gb|U38906|BRU38906 [1353517]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 50 protein links, or 3 nucleotide neighbors)

X91149

Bacteriophage phi-C31 DNA cos region
gil1107473|embl|X91149|APHIC31C [1107473]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 6 protein links, or 1 nucleotide neighbor)

V00642

phage MS2 genome
gil15081|embl|V00642|LEMS2X [15081]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 8 MEDLINE links, 4 protein links, or 20 nucleotide neighbors)

V01146

Genome of bacteriophage T7
gil431187|embl|V01146|T7CG [431187]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 13 MEDLINE links, 60 protein links, 105 nucleotide neighbors, or 1 genome link)

X78401

Bacteriophage P22 right operon, orf 48, replication genes 18 and 12, nin region genes, nluG phosphatase, late control gene 23, orf 60, complete cds, late control region, start of lysis gene 13
gil512343|emb|X78401|POP22NIN [512343]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 13 protein links, or 4 nucleotide neighbors)

Y00408

Bacteriophage T4 gene t for lysis protein
gil15368|embl|Y00408|MYT4T [15368]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

Z26590

5

250

10

Bacteriophage mv4 lysA and lysB genes
gil410500|embl|Z26590|MV4LYSAB [410500]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 4
protein links)

15

X07809

Phage phiX174 lysis (E) gene upstream region
gil15094|embl|X07809|M1PHXE [15094]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 2 protein links, or 4 nucleotide neighbors)

20

Z34528

Lactococcal bacteriophage c2 lysis gene
gil506455|embl|Z34528|LBC2LYSIN [506455]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 1 protein link, or 4 nucleotide neighbors)

25

X15031

Bacteriophage fr RNA genome
gil15071|embl|X15031|LEBFRX [15071]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 4 protein links, 9 nucleotide neighbors, or 1 genome link)

30

X80191

Bacteriophage PP7 mRNA for maturation, coat, lysis and replicase
proteins
gil517237|embl|X80191|BPP7PR [517237]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 4 protein links, or 1 genome link)

35

X85010

Bacteriophage A511 ply511 gene
gil853748|embl|X85010|BPA511PLY [853748]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 3 protein links, or 1 nucleotide neighbor)

40

X85009

Bacteriophage A500 hol500 and ply500 genes
gil853744|embl|X85009|BPA500PLY [853744]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 3 protein links, or 4 nucleotide neighbors)

45

X85008

Bacteriophage A118 hol118 and ply118 genes
gil853740|embl|X85008|BPA118PLY [853740]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 3 protein links, or 1 nucleotide neighbor)

50

55

5

Z35638

251

Bacteriophage phi-X174 genes for lysis protein and beta-lactamase
gil520996|embZ35638|BPLYSPR [520996]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 2 protein links, or 516 nucleotide neighbors)

10

J02459

Bacteriophage lambda, complete genome
gil215104|gbJ02459|LAMCG [215104]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,87 MEDLINE
links, 67 protein links, 190 nucleotide neighbors, or 1 genome link)

15

X87674

Bacteriophage P1 lydA & lydB genes
gil974763|embX87674|BACP1LYD [974763]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 2 protein links, or 2 nucleotide neighbors)

20

X87673

Bacteriophage P1 gene 17
gil974761|embX87673|BACP117 [974761]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 1 protein link, or 1 nucleotide neighbor)

25

M14784

Bacteriophage T3 strain amNG220B right end, tail fiber protein, lysis
protein and DNA packaging proteins, complete cds
gil215810|gbM14784|PT3RE [215810]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 9 protein links, or 10 nucleotide neighbors)

30

M11813

Bacteriophage PZA (from B.subtilis), complete genome
gil216046|gbM11813|PZACG [216046]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE
links, 27 protein links, 17 nucleotide neighbors, or 1 genome link)

35

M16812

Bacteriophage K3 't' lysis gene, complete cds
gil215503|gbM16812|PK3LYST [215503]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 1 protein link, or 4 nucleotide neighbors)

40

J04356

Bacteriophage P22 proteins 15 (complete cds), and 19 (3' end) genes
gil215265|gbJ04356|P2215P [215265]

45

50

55

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 2 nucleotide neighbors)

J04343

Bacteriophage JP34 coat and lysis protein genes, complete cds, and replicase protein gene, 5' end
gi215076|gbJ04343|JF3COLY [215076]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 2 nucleotide neighbors)

J02482

Bacteriophage phi-X174, complete genome
gi216019|gbJ02482|PX1CG [216019]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,23 MEDLINE links, 11 protein links, 26 nucleotide neighbors, or 1 genome link)

M99441

Bacteriophage T4 anti-sigma 70 protein (asiA) gene, complete cds and lysis protein, 3' end
gi215820|gb|M99441|IPT4ASIA [215820]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 2 protein links, or 2 nucleotide neighbors)

M65239

Bacteriophage 21 lysis genes S, R, and Rz, complete cds
gi215466|gb|M65239|PH2LYSGEN [215466]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

M10637

Phage G4 D/E overlapping gene system, encoding D (morphogenetic) and E (lysis) proteins
gi215427|gb|M10637|PG4DE [215427]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 12 nucleotide neighbors)

J02454

Bacteriophage G4, complete genome
gi215415|gbJ02454|PG4CG [215415]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 11 protein links, 20 nucleotide neighbors, or 1 genome link)

J02580

Bacteriophage PA-2 (E.coli porcine strain isolate) Rz gene, 5'end; ORF2, outer membrane porin protein (lc) and ORF1 genes, complete cds
gi215366|gbJ02580|PA2LC [215366]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 4 nucleotide neighbors)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.